

Universidade de São Paulo  
Escola Superior de Agricultura “Luiz de Queiroz”

Análise AMMI e AMMI robusta em dados de trigo com observações  
discrepantes

**Werlleson Nascimento**

Dissertação apresentada para obtenção do título de  
Mestre em Ciências. Área de concentração: Estatística e  
Experimentação Agronômica

Piracicaba  
2023

Werlleson Nascimento  
Agrônomo

Análise AMMI e AMMI robusta em dados de trigo com observações discrepantes

Orientador:  
Prof. Dr. **CARLOS TADEU DIAS DOS SANTOS**

Dissertação apresentada para obtenção do título de  
Mestre em Ciências. Área de concentração: Estatística e  
Experimentação Agronômica

Piracicaba  
2023

**Dados Internacionais de Catalogação na Publicação**  
**DIVISÃO DE BIBLIOTECA – DIBD/ESALQ/USP**

Nascimento, Werlleson

Análise AMMI e AMMI robusta em dados de trigo com observações discrepantes / Werlleson Nascimento. - - Piracicaba, 2023.

68 p.

Dissertação (Mestrado) - - USP / Escola Superior de Agricultura "Luiz de Queiroz".

1. Grupos de experimentos 2. R-AMMI 3. Reamostragem 4. Bootstrap  
5. Permutação I. Título

## DEDICATÓRIA

À Deus  
**Dedico**

## AGRADECIMENTOS

Agradeço à Deus, o ao autor e consumidor de minha fé. Sem Ele, jamais poderia estar no lugar que estou.

Agradeço à minha mãe, Maria Sulamita Nascimento, por todo o suporte, amor e carinho que sempre me deu durante toda a minha vida, inclusive durante este percurso acadêmico.

Aos meus amados irmãos em Cristo Jesus, nosso Senhor, principalmente os da Assembleia de Deus em Santarém e os da Assembleia de Deus Ministério do Belém em Piracicaba, por todo o apoio e orações.

Aos meus amigos que sempre me incentivaram e ficaram felizes com minhas conquistas.

Ao meu orientador, Prof. Dr. Carlos Tadeu dos Santos Dias, por todos os ensinamentos, tanto os acadêmicos, quanto os profissionais.

Aos professores e demais funcionários do Programa de Pós-graduação em Estatística e Experimentação Agronômica (PPGEEA) da Universidade de São Paulo, na Escola Superior de Agricultura “Luiz de Queiroz” (USP/ESALQ), assim como, os demais funcionários da escola que sempre me atenderam com muito carinho.

A todos que, de algum modo, contribuíram para eu chegar onde estou.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001.

## SUMÁRIO

RESUMO .....	7
ABSTRACT .....	8
LISTA DE FIGURAS .....	9
LISTA DE TABELAS .....	11
1. INTRODUÇÃO.....	13
2. REVISÃO DE LITERATURA .....	15
2.1. A cultura do trigo.....	15
2.2. Plantio direto .....	15
2.3. A interação genótipo ambiente .....	15
2.4. Análise de grupos de experimentos .....	16
2.4.1. Análise de variância individual dos experimentos .....	16
2.4.1. Análise de variância conjunta dos experimentos.....	17
2.5. Decomposição em valor singular (DVS).....	18
2.6. O modelo AMMI.....	19
2.6.1. Ajuste dos efeitos principais por anova.....	21
2.6.2. Desdobramento da soma de quadrado da interação por DVS .....	23
2.6.3. Escolha do número de termos multiplicativos.....	24
2.6.3.1. Teste de Gollob.....	25
2.6.3.2. Validação cruzada.....	26
2.7. Limitações do modelo AMMI .....	27
2.7.1. Método baseado em reamostragem (resampling-based) .....	27
2.7.2. R-AMMI.....	29
3. MATERIAL E MÉTODOS.....	31
3.1. Área experimental.....	31
3.2. Desenho experimental e material vegetal.....	31
3.3. Análise estatística .....	32
3.4. Pacotes estatísticos .....	33
4. RESULTADOS E DISCUSSÃO .....	35
4.1. Análise exploratória.....	35
4.2. Análise de resíduos.....	36

4.2.1. Dados completos .....	36
4.2.2. Trigo comum ( <i>Triticum aestivum</i> ) .....	40
4.2.3. Trigo durum ( <i>Triticum turgidum</i> ) .....	44
4.3. Análise AMMI dos dados transformados de trigo comum ( <i>T. aestivum</i> ) .....	48
4.4. Aplicação dos modelos AMMI usando reamostragem .....	51
4.4.1. Dados completos .....	51
4.4.2. Trigo durum ( <i>T. turgidum</i> ) .....	53
4.5. Aplicando o modelo R-AMMI .....	55
4.5.1. Dados completos .....	55
4.5.2. Trigo durum ( <i>T. turgidum</i> ) .....	57
4.6. Considerações sobre os resultados das análises .....	59
5. CONCLUSÕES .....	63
REFERÊNCIAS .....	65

## RESUMO

### **Análise AMMI e AMMI robusta em dados de trigo com observações discrepantes**

Ensaio agrícolas implantados em diferentes locais são bastante comuns, principalmente em fases finais de programas de melhoramento genético. Estudar a adaptabilidade e a estabilidade dos materiais genéticos é de grande interesse para o melhorista, seja para identificar genótipos superiores, seja para fazer recomendações regionalizadas. Estudar o efeito do ambiente sobre o rendimento das culturas é um dos principais desafios deste tipo de pesquisa, considerando-se que a interação genótipo  $\times$  ambiente pode se tornar bastante complexa. O tipo de análise mais comum para situações como essa é fazer a análise de variância de grupos de experimentos e aplicação de testes de comparação de médias. No entanto, também é possível fazer uso de métodos estatísticos multivariados. Entre as metodologias mais proeminentes, destaca-se o modelo AMMI. Esta metodologia combina em uma única análise a ANOVA, para estudar os efeitos aditivos do modelo, e a Análise de Componentes Principais, ou Decomposição em Valor Singular para decompor a interação. Ocorre que esta técnica possui pressupostos muito fortes, como a necessidade de normalidade e homogeneidade de variâncias residual. Para contornar esta limitação, alternativas têm sido propostas, entre estas destaca-se métodos de reamostragem e o algoritmo R-AMMI. Assim, objetivou-se neste trabalho aplicar o modelo AMMI tradicional e os métodos de reamostragem e algoritmo R-AMMI quando as pressuposições do modelo não forem atendidas, em um conjunto de dados proveniente da estação experimental do CIMMYT (Centro Internacional de Mejoramiento de Maíz y Trigo), localizada em Ciudad Obregon, Sonora, no noroeste do México, de um ensaio realizado em quatro ambientes distintos: (E1) canteiros convencionalmente arados com irrigação total, (E2) canteiros convencionalmente arados com irrigação reduzida, (E3) plantio direto com irrigação total e (E4) plantio direto com irrigação reduzida. Dentro de cada ambiente foi implantado um delineamento em blocos completamente casualizados, com três blocos e 24 genótipos de duas espécies de trigo, sendo 12 genótipos de trigo comum (*Triticum aestivum*) e 12 de trigo durum (*Triticum turgidum*). A variável resposta analisada neste estudo foi o rendimento de grãos (kg/ha), referentes à safra de 2017-2018. Os ambientes E1 e E4 foram bastante similares entre si, indicando que é possível ter um ambiente altamente produtivo com práticas conservacionistas do solo. O sistema de plantio direto pode ser prejudicado no caso de excesso de irrigação, possivelmente devido a maior infiltração de água nesses ambientes, que podem resultar estresse hídrico às plantas. Os algoritmos R-AMMI e o método baseado em reamostragem são importantes formas para contornar a situação de se trabalhar com dados contaminados com observações discrepantes. Como sugestão para estudos futuros, recomenda-se combinar o algoritmo de reamostragem com o R-AMMI, aplicando-o diretamente nos resíduos da regressão robusta.

Palavras-chave: Grupo de experimentos, R-AMMI, Reamostragem, Bootstrap, Permutação



## ABSTRACT

**Robust AMMI and AMMI analysis on wheat data with outliers**

Agricultural trials implemented in different locations are quite common, especially in the final stages of genetic improvement programs. Studying the adaptability and stability of genetic materials is of great interest to breeders, either to identify superior genotypes or to make regionalized recommendations. Studying the effect of the environment on crop yields is one of the main challenges of this type of research, considering that the genotype  $\times$  environment interaction can become quite complex. The most common type of analysis for situations like this is to perform the analysis of variance of groups of experiments and application of mean comparison tests. However, it is also possible to use multivariate statistical methods. Among the most prominent methodologies, the AMMI model stands out. This methodology combines in a single analysis the ANOVA to study the additive effects of the model and Principal Component Analysis or Singular Value Decomposition to decompose the interaction. However, this technique has very strong assumptions, such as the need for normality and homogeneity of residual variances. To overcome this limitation, alternatives have been proposed, including resampling methods and the R-AMMI algorithm. Thus, the objective of this work was to apply the traditional AMMI model and resampling methods and R-AMMI algorithm when the model assumptions are not met, in a dataset from the experimental station of CIMMYT (Centro Internacional de Mejoramiento de Maíz y Trigo) located in Ciudad Obregon, Sonora, in northwest Mexico, from a trial conducted in four different environments: (E1) conventionally plowed beds with full irrigation, (E2) conventionally plowed beds with reduced irrigation, (E3) no-till with full irrigation and (E4) no-till with reduced irrigation. Within each environment, a completely randomized block design was implanted, with three blocks and 24 genotypes of two species of wheat, 12 of common wheat (*Triticum aestivum*) and 12 of durum wheat (*Triticum turgidum*). The response variable analyzed in this study was grain yield (kg/ha) for the 2017-2018 crop. Environments E1 and E4 were quite similar, indicating that it is possible to have a highly productive environment with soil conservation practices. The no-till system can be impaired in case of excessive irrigation, possibly due to greater water infiltration in these environments, which can result in water stress to plants. The R-AMMI algorithm and the resampling-based method are important ways to deal with the situation of working with data contaminated with peripheral observations. As a suggestion for future studies, it is recommended to combine the resampling algorithm with R-AMMI, applying it directly to the residuals of robust regression.

Keywords: Group of experiment, R-AMMI, Resampling, Bootstrap, Permutation

## LISTA DE FIGURAS

Figura 1. Rendimento de grãos de trigo (kg/ha) por genótipos segundo os ambientes.....	35
Figura 2. Rendimento de grãos de trigo (kg/ha) por ambientes segundo os genótipos.....	36
Figura 3. Gráficos de genótipos <i>versus</i> resíduos estudentizados (coluna à esquerda) e quantis normais <i>versus</i> quantis observados na amostra (coluna à direita) para cada ambiente, considerando-se os dados completos. ....	37
Figura 4. Gráficos de resíduos dos dados completos para ambientes, genótipos e combinação genótipo e ambiente.....	38
Figura 5. Resíduos estudentizados <i>versus</i> valores esperados.....	39
Figura 6. Gráficos de resíduos dos dados completos após transformação para ambientes, genótipos e combinação genótipo e ambiente.....	39
Figura 7. Gráficos de genótipos <i>versus</i> resíduos estudentizados (coluna à esquerda) e quantis normais <i>versus</i> quantis observados na amostra (coluna à direita) para cada ambiente, considerando-se os dados de trigo comum ( <i>Triticum aestivum</i> ).....	41
Figura 8. Gráficos de resíduos relacionados aos dados de <i>T. aestivum</i> para ambientes, genótipos e combinação genótipo e ambiente.....	42
Figura 9. Gráficos de resíduos relacionados aos dados transformados de trigo comum ( <i>T. aestivum</i> ) para ambientes, genótipos e combinação genótipo e ambiente. ....	43
Figura 10. Gráficos de genótipos <i>versus</i> resíduos estudentizados (coluna à esquerda) e quantis normais <i>versus</i> quantis observados na amostra (coluna à direita) para cada ambiente, considerando-se os dados de trigo durum ( <i>Triticum turgidum</i> ).....	45
Figura 11. Gráficos de resíduos relacionados aos dados de trigo durum ( <i>T. turgidum</i> ) para ambientes, genótipos e combinação genótipo e ambiente.....	46
Figura 12. Gráficos de resíduos relacionados aos dados de trigo durum ( <i>T. turgidum</i> ) após transformação por box-cox para ambientes, genótipos e combinação genótipo e ambiente. ....	47
Figura 13. Raíz quadrada do quadrado médio da diferença de predição para os dados transformados de trigo comum ( <i>T. aestivum</i> ) e ajustados para a escala original dos dados. ....	49
Figura 14. Biplot AMMI1 (IPC1) <i>versus</i> rendimento médio (kg/ha) de ambientes e genótipos. ....	50
Figura 15. Biplot AMMI2 (IPC1 <i>versus</i> IPC2) dos dados transformados de trigo comum ( <i>T. aestivum</i> ).....	51
Figura 16. Biplot AMMI1 (IPC1) <i>versus</i> rendimento médio (kg/ha) dos dados completos de trigo, considerando-se as espécies <i>T. aestivum</i> e <i>T. durum</i> . ....	52
Figura 17. Biplot AMMI2 (IPC1 <i>versus</i> IPC2) para os dados completos, considerando-se as espécies <i>T. aestivum</i> e <i>T. Durum</i> .....	53
Figura 18. Biplot AMMI1 (IPC1) <i>versus</i> rendimento médio (kg/ha) para os dados de trigo durum ( <i>T. turgidum</i> ). ....	54
Figura 19. Biplot AMMI2 (IPC1 <i>versus</i> IPC2) dos dados de trigo durum ( <i>T. turgidum</i> ). ....	55
Figura 20. Biplot AMMI1 (IPC1) <i>versus</i> rendimento médio (kg/ha) dos dados completos utilizando o algoritmo R-AMMI.....	56

Figura 21. Biplot AMMI2 (IPC1 <i>versus</i> IPC2) dos dados completos usando o algoritmo R-AMMI. ....	57
Figura 22. Biplot AMMI1 (IPC1 <i>versus</i> rendimento médio (kg/ha) para a espécie de trigo durum ( <i>T. turgidum</i> ) usando o algoritmo R-AMMI.....	58
Figura 23. Biplot AMMI2 (IPC1 <i>versus</i> IPC2) para a espécie de trigo durum ( <i>T. turgidum</i> ) usando o algoritmo R-AMMI.....	59

## LISTA DE TABELAS

Tabela 1. Esquema da análise de variância em dab para os experimentos individuais. ....	17
Tabela 2. Esquema da análise de variância conjunta em dab. ....	18
Tabela 3. Esquema da análise de variância considerando-se um delineamento inteiramente aleatorizado segundo o sistema proposto por gollob.....	26
Tabela 4. Genótipos de trigo comum ( <i>Triticum aestivum</i> ) e trigo durum ( <i>Triticum turgidum</i> ).....	32
Tabela 5. Testes de normalidade e homogeneidade de variâncias para a variável rendimento de grãos (kg/ha) para os dados completos. ....	38
Tabela 6. Testes de homogeneidade de variâncias para cada fonte de variação e normalidade dos resíduos. ....	38
Tabela 7. Testes de homogeneidade de variâncias para cada fonte de variação e normalidade dos resíduos ....	40
Tabela 8. Testes de normalidade e homogeneidade de variâncias para a variável rendimento de grãos (kg/ha) para os dados de trigo comum ( <i>Triticum aestivum</i> ).....	42
Tabela 9. Testes de homogeneidade de variâncias para cada fonte de variação e normalidade dos resíduos relacionados a espécie <i>T. aestivum</i> . ....	43
Tabela 10. Testes de homogeneidade de variâncias para a variável rendimento de grãos (kg/ha) transformada, para cada fonte de variação e normalidade dos resíduos para a espécie <i>T. aestivum</i> . ....	43
Tabela 11. Testes de normalidade e homogeneidade de variâncias para a variável rendimento de grãos (kg/ha), considerando-se os dados de trigo durum ( <i>T. turgidum</i> ). ....	46
Tabela 12. Resumo da análise de variância para a variável rendimento de grãos (kg/ha), por locais, considerando-se os genótipos de trigo durum ( <i>T. turgidum</i> ).....	46
Tabela 13. Testes de homogeneidade de variâncias para a variável rendimento de grãos (kg/ha), para cada fonte de variação e normalidade dos resíduos, considerando-se os dados de trigo durum ( <i>T. turgidum</i> ). ....	47
Tabela 14. Testes de homogeneidade de variâncias para a variável rendimento de grãos (kg/ha) transformada, para cada fonte de variação e normalidade dos resíduos, considerando-se os dados de trigo durum ( <i>T. turgidum</i> ). ....	48
Tabela 15. Análise de variância para os dados transformados de rendimento de grãos em kg/ha de trigo comum ( <i>T. aestivum</i> ).....	48
Tabela 16. Estatística <b>Tr</b> e valores-p calculados utilizando o método de <i>bootstrap</i> não-paramétrico e permutação para os dados completos.....	51
Tabela 17. Estatística <b>Tr</b> e valores-p calculados utilizando o método de <i>bootstrap</i> não-paramétrico e permutação para os dados de trigo .....	53



## 1. INTRODUÇÃO

Em programas de melhoramento genético é comum a instalação de experimentos em diferentes locais ou ambientes, com o objetivo de estudar o desempenho agrônômico de culturas, assim como, estudar o efeito de interação genótipo  $\times$  ambiente (*GEI – genotype, environment interaction*) sobre as plantas, principalmente em suas fases finais. Nesta etapa, o fator ambiental será o que mais influenciará a expressão fenotípica, considerando-se que o efeito de genótipo é mais estático (SCAVO; MAUROMICALE; IERNA, 2023). Compreender este efeito permite ao melhorista identificar genótipos superiores, estáveis e com ampla capacidade adaptativa, ou mesmo fazer recomendações mais regionalizadas, considerando-se o desempenho das culturas em campo.

Estes ensaios multiambientais envolvendo grupos de experimentos podem ser analisados usando uma análise de variância (ANOVA) apropriada, considerando-se os efeitos principais de genótipos, ambientes e GEI, podendo-se aplicar testes de comparações de médias em seguida. No entanto, situações como essa também podem ser analisadas usando métodos estatísticos multivariados. Entre os métodos mais conhecidos, destaca-se o modelo AMMI. Esta técnica combina em uma única análise a ANOVA e Análise de Componentes Principais (ACP) ou Decomposição em Valor Singular (DVS), com o intuito de estudar simultaneamente os termos aditivos e multiplicativos (GEI) do modelo, no caso de GEI significativa (GAUCH, 2013).

Uma das limitações do modelo AMMI é a necessidade de atendimento de pressupostos da ANOVA, como a normalidade de resíduos e homoscedasticidade de variâncias residuais, o que pode tornar inviável a sua aplicação quando os ambientes são muito heterogêneos, em função de não ser possível escolher adequadamente o número de termos multiplicativos significativos do modelo. Contudo, algoritmos robustos têm sido propostos para lidar com este tipo de situação. Um algoritmo baseado em reamostragem proposto por Malik, Forkman e Piepho (2019) é uma das alternativas para contornar as limitações do modelo AMMI tradicional, em função de usar *bootstrap* não paramétrico e permutação, não sendo necessário assumir nenhuma distribuição teórica para o conjunto de dados. Além disso, é possível aplicar também um modelo AMMI robusto, denominado de R-AMMI, proposto por Rodrigues, Monteiro e Lourenço (2016). O modelo é uma generalização da abordagem clássica, onde métodos estatísticos robustos são utilizados para substituir a ANOVA e a DVS.

Portanto, o objetivo deste trabalho foi aplicar o modelo AMMI em um conjunto de dados que atende as pressuposições da análise, assim como, analisar outros dois conjuntos de

dados que não atendem aos pressupostos usando a combinação do algoritmo de reamostragem com o AMMI e o algoritmo R-AMMI. Os dados deste trabalho são provenientes de um ensaio de 24 genótipos de duas espécies de trigo (*Triticum aestivum* e *Triticum durum*), avaliados em quatro ambientes agronômicos, na estação experimental do CIMMYT (*Centro Internacional de Mejoramiento de Maíz y Trigo*), localizada em Ciudad Obregon, Sonora, no noroeste do México, onde avaliou-se o rendimento de grãos (kg/ha).

## 2. REVISÃO DE LITERATURA

### 2.1. A cultura do trigo

O trigo é a cultura mais cultivada no mundo com mais de 221 milhões de hectares (ha) plantados nos anos safras de 2021/2022, e uma produção superior a 779 milhões de toneladas (USDA, 2023). Além disso, a produção global aumentou em 1,3% entre 2000 e 2019, tornando-a de grande importância para a segurança alimentar global (FISCHER et al., 2022). As plantas de trigo pertencem à família Poaceae, em cinco espécies: *Triticum einkorn*, *Triticum emmer*, *Triticum durum*, *Triticum aestivum*, e *Triticum tritcale*. Este último é resultante de um cruzamento artificial de *T. durum* e centeio. A espécie mais cultivada é *T. aestivum*, seguida por *T. durum*, sendo os grãos e a farinha de trigo o ingrediente base para diversos produtos (WIESER; KOEHLER; SCHERF, 2020).

### 2.2. Plantio direto

O sistema de plantio direto, ou cultivo mínimo, é uma técnica de cultivo conservacionista do solo que consiste em revolvê-lo minimamente, bem como, acumular matéria orgânica em sua superfície, proveniente da cultura anterior que estava presente na área produtiva. Nesse sistema, as sementes da nova cultura são semeadas diretamente nas palhas. Esta forma de manejo ajuda a manter a produtividade do solo, assim como os estoques de água e rendimento de grãos, previne erosão devido ao não impacto direto da chuva no solo, aumenta os estoques de carbono, nitrogênio e a atividade microbiana, além de reduzir o uso de mão-de-obra e energia, em função de não ser mais necessário, por exemplo, arar a terra (EL HARCHE et al., 2023; LIU et al., 2022; QI et al., 2023; YANG et al., 2023).

### 2.3. A interação genótipo ambiente

O progresso no melhoramento é dependente da identificação de genótipos que sejam mais adaptados ao ambiente em que estão sendo estudados (CHENU et al., 2011). Assim, o estudo da interação genótipo × ambiente é de grande importância em programas de melhoramento genético, particularmente em suas fases finais, considerando-se que o desempenho agrônomico das culturas é altamente influenciado por variações ambientais. Portanto, buscar-se genótipos estáveis e amplamente adaptáveis à diferentes condições ambientais é extremamente



vantajoso para o melhorista, além de ser possível obter informações relevantes relacionadas à interação ao se identificar ambientes que tragam melhorias em características fenotípicas das plantas (SCAVO; MAUROMICALE; IERNA, 2023).

## 2.4. Análise de grupos de experimentos

A implantação de experimentos em diferentes ambientes é muito comum em programas de melhoramento genético porque permite ao melhorista avaliar o desempenho agrônômico de genótipos sob diferentes condições ambientais. Normalmente esses experimentos utilizam delineamentos mais simples, como o inteiramente aleatorizado (DIA) ou delineamento aleatorizado em blocos (DAB). Trata-se, portanto, da repetição do mesmo experimento, contendo os mesmos tratamentos e delineamento, variando-se apenas o local. Antes da realização da análise conjunta, é realizada uma análise individual dos experimentos por ambiente, bem como, é avaliado se os quadrados médios residuais são homogêneos, para só então, prosseguir-se com a análise conjunta.

### 2.4.1. Análise de variância individual dos experimentos

A análise de variância é uma poderosa técnica estatística, proposta por Fisher (1919), capaz de particionar a variação presente no conjunto de dados em efeitos principais e de interação, considerando-se uma ANOVA de duas vias, por exemplo. Em ensaios de rendimento de cultivares, é mais comum que efeitos principais de ambientes sejam maiores, seguidos da interação e, por fim, do efeito de genótipos (GAUCH, 2006). Um modelo envolvendo os  $g$  genótipos ( $i = 1, 2, \dots, g$ ) com  $r$  blocos ( $k = 1, 2, \dots, r$ ), pode ser representado como (1):

$$y_{ik} = \mu + g_i + \beta_k + \varepsilon_{ik} \quad (1)$$

em que:

$y_{ik}$ : é o valor observado no  $i$ -ésimo genótipo, na  $k$ -ésima repetição com  $i = 1, \dots, g$ , e  $k = 1, 2, \dots, r$ ;

$\mu$ : é uma constante comum às observações;

$g_i$ : é o efeito fixo do  $i$ -ésimo genótipo;

$\beta_k$ : é o efeito fixo do  $k$ -ésimo bloco;

$\varepsilon_{ik}$ : é o erro experimental associado às observações, assumindo  $\varepsilon_{ik} \sim N(0, \sigma^2)$  e  $cov(\varepsilon_{ik}, \varepsilon_{i'k'}) = 0$  para  $(i, k) \neq (i', k')$ .

O esquema da ANOVA é apresentado na Tabela 1. Os graus de liberdade do resíduo foram obtidos a partir da diferença entre os graus de liberdade total e a soma dos graus de liberdade de genótipos e blocos:

**Tabela 1.** Esquema da análise de variância em DAB para os experimentos individuais.

Fontes de Variação	Graus de Liberdade (dados completos)
Genótipos (G)	$(g - 1)$
Blocos (B)	$(r - 1)$
Resíduo	$(g - 1)(r - 1)$
Total	$gr - 1$

#### 2.4.1. Análise de variância conjunta dos experimentos

O modelo da análise de variância conjunta para cada experimento envolvendo os  $e$  ambientes ( $j = 1, 2, \dots, e$ ) e os  $g$  genótipos ( $i = 1, 2, \dots, g$ ) com  $r$  blocos ( $k = 1, 2, \dots, r$ ), pode ser representado como (2):

$$y_{ijk} = \mu + g_i + e_j + (ge)_{ij} + \beta_{k(j)} + \varepsilon_{ijk} \quad (2)$$

em que:

$y_{ijk}$ : é o valor observado no  $i$ -ésimo genótipo do  $j$ -ésimo ambiente e na  $k$ -ésima repetição com  $i = 1, \dots, g$ ,  $j = 1, \dots, e$  e  $k = 1, 2, \dots, r$ ;

$\mu$ : é uma constante comum às observações;

$g_i$ : é o efeito fixo do  $i$ -ésimo genótipo;

$e_j$ : é o efeito fixo do  $j$ -ésimo ambiente;

$\beta_{k(j)}$ : é o efeito fixo do  $k$ -ésimo bloco dentro do  $j$ -ésimo ambiente;

$(ge)_{ij}$ : é o efeito fixo da interação entre o  $i$ -ésimo genótipo e o  $j$ -ésimo ambiente;

$\varepsilon_{ijk}$ : é o erro experimental associado às observações, assumindo  $\varepsilon_{ijk} \sim N(0, \sigma^2)$  e  $cov(\varepsilon_{ijk}, \varepsilon_{i'j'k'}) = 0$  para  $(i, j, k) \neq (i', j', k')$ .

O esquema da ANOVA é apresentado na Tabela 2. Os graus de liberdade do resíduo foram obtidos a partir da diferença entre os graus de liberdade total e a soma dos graus de liberdade das demais fontes de variação (genótipos, ambientes e interação):

**Tabela 2.** Esquema da análise de variância conjunta em DAB.

Fontes de Variação	Graus de Liberdade (dados completos)
Genótipos (G)	$(g - 1)$
Ambientes (E)	$(e - 1)$
Blocos d. Ambientes (B[E])	$e(r - 1)$
Interação (GE)	$(g - 1)(e - 1)$
Resíduo	$e(g - 1)(r - 1)$
Total	$ger - 1$

## 2.5. Decomposição em valor singular (DVS)

A DVS foi proposta originalmente por Karl Pearson (1901) e consiste em fatorar uma matriz original qualquer  $\mathbf{X}_{m \times p}$  com  $m \geq p$ , em três submatrizes, sendo uma contendo os autovalores e as outras duas são matrizes ortogonais que armazenam os autovetores, correspondentes à cada autovalor, e estão associados às linhas e colunas da matriz  $\mathbf{X}$ . Os autovalores são números reais, não negativos, posicionados em ordem decrescente que indicam a importância relativa das linhas e colunas da matriz original. Isto é bastante desejável, considerando-se que em pesquisas agrônômicas sempre se está trabalhando com a projeção de muitos pontos no espaço, buscando-se o melhor ajuste possível, podendo-se obter uma aproximação para uma matriz qualquer.

A DVS origina novos eixos, também chamados de componentes principais (PC), onde o primeiro PC retém a maior parte da variação dos dados possível, correspondendo a reta de mínimos quadrados através do espaço gerado originalmente pelos pontos. Esta ideia se aplica de igual modo ao segundo componente, onde os dois primeiros PCs formam o plano de mínimos quadrados. Como os primeiros componentes são capazes de reter a maior parte da variação presente no conjunto de dados, é possível representar a maior parte da informação em alguns poucos componentes, reduzindo-se expressivamente a dimensionalidade dos dados.

A decomposição em valor singular, portanto, é uma generalização da decomposição espectral. Enquanto esta última se limita a matrizes quadradas e simétricas, a DVS pode ser utilizada para construir uma aproximação para uma matriz real de uma ordem qualquer. A fatoração de uma matriz  $\mathbf{X}_{m \times p}$  em um produto de três matrizes, considerando-se  $m \geq p$ , pode ser expressa como (3):

$$\mathbf{X} = \mathbf{U}\mathbf{D}\mathbf{V}^T \quad (3)$$

Em que  $\mathbf{U}^T\mathbf{U} = \mathbf{V}^T\mathbf{V} = \mathbf{V}\mathbf{V}^T = \mathbf{I}_p$ . A matriz  $\mathbf{U}$  armazena os  $p$  autovetores ortonormalizados associados com os  $p$  maiores autovalores de  $\mathbf{X}\mathbf{X}^T$  e  $\mathbf{V}$  os  $p$  autovetores ortonormalizados associados com os  $p$  maiores autovalores de  $\mathbf{X}^T\mathbf{X}$ . As matrizes  $\mathbf{U}$  e  $\mathbf{V}^T$  contém os autovetores associados as linhas e as colunas da matriz  $\mathbf{X}$ , respectivamente.  $\mathbf{D}$  é uma matriz diagonal que contém os valores singulares não nulos, ou seja, a raiz quadrada dos autovalores de  $\mathbf{X}^T\mathbf{X}$  ou de  $\mathbf{X}\mathbf{X}^T$  com  $(\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0)$  (LIMA NETO; RODRIGUES, 2023).

De outra maneira, a DVS pode ser escrita como um somatório  $\sum_{k=1}^p \lambda_k \mathbf{u}_k \mathbf{v}_k^T$  em que  $\lambda_k$  é o  $k$ -ésimo autovalor de  $\mathbf{X}$ , com  $\mathbf{u}_k \mathbf{v}_k^T$  sendo vetores singulares (coluna e linha). Assim, fazendo-se um somatório de  $n$  termos, inferiores a  $p$ , é possível obter uma aproximação de posto  $n$  para a matriz  $\mathbf{X}$ . Esta nova matriz é a melhor aproximação de posto  $n$  para a matriz original  $\mathbf{X}$  e com a menor soma de quadrados dos desvios dos elementos desta matriz para os elementos correspondentes na matriz resultantes (GOOD, 1969). Os vetores  $\mathbf{u}_k \mathbf{v}_k^T$  podem ser arranjados em duas matrizes  $\mathbf{U}$  e  $\mathbf{V}$ , sendo que as colunas de  $\mathbf{U}$  são compostas de cada vetor  $\mathbf{u}_k$  e as linhas de  $\mathbf{V}$  são compostas por cada vetor  $\mathbf{v}_k^T$ . Além disso, se arranjarmos cada autovalor em uma matriz  $\mathbf{S}$  com  $(\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0)$ , chegamos novamente na mesma expressão indicada anteriormente:  $\mathbf{X} = \mathbf{U}\mathbf{S}\mathbf{V}^T$

Na análise AMMI, a matriz  $\mathbf{X}$  será representada pela matriz de interações  $\mathbf{GE}$ . Neste sentido, o objetivo de se aplicar a DVS reside no fato de se desejar obter uma aproximação para a matriz  $\mathbf{GE}$  por outra matriz de posto inferior, para se tentar compreender melhor a estrutura adjacente aos dados desta matriz (DUARTE; VENCOSKY, 1999).

## 2.6. O modelo AMMI

Para estudar o comportamento diferencial das culturas em campo, técnicas estatísticas têm sido propostas. Entre estas, destaca-se o Modelo de Efeito Principal Aditivo e Interação Multiplicativa ou, simplesmente, modelo AMMI (Additive Main Effects And Multiplicative Interaction). O modelo aplica a análise de variância (ANOVA) para particionar os efeitos principais de genótipo ( $G$ ), ambiente ( $E$ ) e da interação ( $GE$ ), semelhante ao modo como é feito em uma análise de grupos de experimentos, e a decomposição em valor singular (DVS) ou Análise de componentes principais (ACP) para decompor a interação. Apesar de o modelo possuir aplicações diversas, o seu uso é mais comum na avaliação de ensaios de rendimento de

plantas em diferentes ambientes, com o objetivo de estudar interações complexas de genótipo  $\times$  ambiente, assim como, aumentar a precisão preditiva (GAUCH, 2013).

De outra maneira, o modelo AMMI é uma técnica uni-multivariada que combina a análise de variância (ANOVA) e a decomposição em valor singular (DVS) ou a análise de componentes principais (ACP) em um único modelo. Tanto a DVS quanto a ACP são técnicas correlatas, porém, possuem uma ligeira diferença em relação aos cálculos relacionados à decomposição da interação. Isto é, por DVS os parâmetros multiplicativos são estimados simultaneamente a partir da matriz  $\mathbf{GE}$  (tratada mais adiante), enquanto por ACP as análises são feitas pelos produtos de  $(\mathbf{GE})(\mathbf{GE})^T$  e  $(\mathbf{GE})^T(\mathbf{GE})$ , o que além de aumentar a quantidade de análises exigirá mais atenção na interpretação dos resultados.

O ajuste do modelo é feito em duas etapas. na etapa inicial, a anova é aplicada somente aos termos aditivos do modelo (médias de genótipos e ambientes). Os resíduos deste modelo, denominados de resíduos de não aditividade do efeito atribuído à interação ( $G \times E$ ), que é a parte multiplicativa do modelo, são armazenados em uma matriz  $\mathbf{GE}_{g \times e} = [(\widehat{ge})_{ij}]$ , chamada de matriz de interações. Na segunda etapa da análise, a matriz  $\mathbf{GE}$  é decomposta via DVS (decomposição em valor singular) ou por ACP. O modelo pode ser escrito, inicialmente, da seguinte forma (4):

$$Y_{ij} = \mu + g_i + e_j + \sum_{k=1}^p \lambda_k \gamma_{ik} \alpha_{jk} + \varepsilon_{ij} \quad (4)$$

Neste modelo, o componente de interação  $(ge)_{ij}$  ganha uma nova abordagem e passa a ser modelado por  $\sum_{k=1}^p \lambda_k \gamma_{ik} \alpha_{jk}$ , via DVS. Pode-se ainda escrever este componente da seguinte forma:  $\sum_{k=1}^p \lambda_k \gamma_{ik} \alpha_{jk} = \sum_{k=1}^n \lambda_k \gamma_{ik} \alpha_{jk} + \sum_{k=n+1}^p \lambda_k \gamma_{ik} \alpha_{jk}$ , com  $n < p$ , em que  $\sum_{k=n+1}^p \lambda_k \gamma_{ik} \alpha_{jk} = \rho_{ij}$  é um resíduo de interação que é descartado. Sendo assim, a resposta média do genótipo  $i$  no ambiente  $j$  é dada por (5):

$$Y_{ij} = \mu + g_i + e_j + \sum_{k=1}^n \lambda_k \gamma_{ik} \alpha_{jk} + \rho_{ij} + \varepsilon_{ij} \quad (5)$$

em que:

$y_{ij}$ : é a resposta média do  $i$ -ésimo genótipo no  $j$ -ésimo ambiente com  $i = 1, \dots, g$  e  $j = 1, \dots, e$ ;

$\mu$ : é uma constante comum a todas as observações, geralmente a média geral;

$g_i$ : é o efeito do  $i$ -ésimo genótipo;

$e_j$ : é o efeito do  $j$ -ésimo ambiente;

$\lambda_k$ : é o  $k$ -ésimo valor singular de  $\mathbf{GE}$  (escalar), considerando-se a DVS, ou ainda, a raiz quadrada do  $k$ -ésimo autovalor das matrizes  $(\mathbf{GE})(\mathbf{GE})^T$  e  $(\mathbf{GE})^T(\mathbf{GE})$  de iguais valores não nulos ( $\lambda_k^2$  é o  $k$ -ésimo autovalor);

$\gamma_{ik}$ :  $i$ -ésimo elemento do  $k$ -ésimo autovetor de genótipos provenientes de  $(\mathbf{GE})(\mathbf{GE})^T$  associado a  $\lambda_k^2$  (ACP);

$\alpha_{jk}$ :  $j$ -ésimo elemento do  $k$ -ésimo autovetor de ambientes  $(\mathbf{GE})^T(\mathbf{GE})$  associado a  $\lambda_k^2$  (ACP);

$\rho_{ij}$ : ruído presente nos dados que permanece se todos os  $k$ -PCAs não são considerados, onde  $k = (1, 2, \dots, p)$  e  $p = \min(g - 1, e - 1)$ , em que  $p$  é o posto de  $\mathbf{GE}$ .

$\varepsilon_{ij}$ : é o erro experimental médio, assumindo  $\varepsilon_{ij} \sim N\left(0, \frac{\sigma^2}{m}\right)$ , em que  $m$  é o número de repetições;

$n$ : número de componentes retidos no modelo ( $n < p$ ).

Uma aproximação de mínimos quadrados para a matriz  $\mathbf{GE}$  pelos  $n$  primeiros termos da DVS é realizada tomando o índice  $k$  até  $n$  no somatório. Considerando-se que o modelo AMMI não busca recuperar toda a soma de quadrado da interação ( $SQ_{g \times e}$ ), mas apenas a parte da informação que é influenciada principalmente pelos genótipos e os ambientes (linhas e colunas de  $\mathbf{GE}$ ), ou seja, separar padrão e ruído, um resíduo adicional  $\rho_{ij}$  é desconsiderado no componente relativo à interação. No entanto, quando  $n = p$ , se tem a decomposição exata da matriz, resultando em um  $\rho_{ij}$  nulo. Sob as restrições de identificabilidade dadas por:  $\sum_i g_i = \sum_j e_j = \sum_i (ge)_{ij} = \sum_j (ge)_j = \sum_{i,j} (ge)_{ij} = 0$ , a DVS da matriz de interações dá origem ao termo  $\sum_{k=1}^n \lambda_k \gamma_{jk} \alpha_{ik} + \rho_{ij}$ . Ademais, assim como na ACP, cada componente retido no modelo capta quantidade de informação cada vez menor da variação presente na matriz de interações ( $\lambda^2_1 \geq \lambda^2_2 \geq \dots \geq \lambda^2_p$ ).

### 2.6.1. Ajuste dos efeitos principais por anova

Na etapa inicial da análise AMMI é feita a uma análise de variância ajustada somente aos efeitos principais do modelo a partir da tabela contendo as médias dos genótipos em cada

ambiente ( $\mathbf{Y}_{g \times e}$ ). Assim, a resposta média de  $r$ -repetições ou blocos do  $i$ -ésimo genótipo no  $j$ -ésimo ambiente é dada por (6):

$$Y_{ij} = \mu + g_i + e_j + \varepsilon_{ij} \quad (6)$$

Neste modelo  $\varepsilon_{ij}$  é o resíduo de não aditividade equivalente ao termo da interação  $(\widehat{ge})_{ij}$ . Assim, sob as restrições de identificabilidade  $\sum_{i=1}^g g_i = \sum_{j=1}^e e_j = 0$ , as soluções de mínimos quadrados ordinários são:

$$\hat{\mu} = \bar{Y}_{..}; \quad \hat{g}_i = \bar{Y}_{i.} - \bar{Y}_{..}; \quad \hat{e}_j = \bar{Y}_{.j} - \bar{Y}_{..};$$

$$\text{Considerando-se: } \bar{Y}_{..} = \frac{\sum_{i=1}^g \sum_{j=1}^e Y_{ij}}{ge}; \quad \bar{Y}_{i.} = \frac{\sum_{j=1}^e Y_{ij}}{e} \text{ e } \bar{Y}_{.j} = \frac{\sum_{i=1}^g Y_{ij}}{g}$$

Com isso, obtém-se uma aproximação de mínimos quadrados para  $\hat{Y}_{ij}$  e seu respectivo resíduo, que corresponde ao termo geral da interação  $(\widehat{ge})_{ij}$ , ambos invariantes:

$$\hat{Y}_{ij} = \bar{Y}_{i.} + \bar{Y}_{.j} - \bar{Y}_{..} \text{ e}$$

$$\varepsilon_{ij} = Y_{ij} - \hat{Y}_{ij} \Leftrightarrow (\widehat{ge})_{ij} = Y_{ij} - \bar{Y}_{i.} - \bar{Y}_{.j} + \bar{Y}_{..}$$

em que:

$Y_{ij}$ : é a resposta média do  $i$ -ésimo genótipo no  $j$ -ésimo ambiente;

$\bar{Y}_{i.}$ : é a média do  $i$ -ésimo genótipo;

$\bar{Y}_{.j}$ : é a média do  $j$ -ésimo ambiente;

$\bar{Y}_{..}$ : é a média geral do experimento.

Considerando-se este último resultado, é possível construir a matriz de interações  $\mathbf{GE}_{g \times e}$  (mostrada a seguir), que será utilizada na segunda etapa da análise.

$$\mathbf{GE} = \begin{bmatrix} (\widehat{ge})_{11} & (\widehat{ge})_{12} & \dots & (\widehat{ge})_{1e} \\ (\widehat{ge})_{21} & (\widehat{ge})_{22} & \dots & (\widehat{ge})_{2e} \\ \vdots & \vdots & \ddots & \vdots \\ [(\widehat{ge})_{g1} & (\widehat{ge})_{g2} & \dots & (\widehat{ge})_{ge}] \end{bmatrix}$$

A partir das médias de cada genótipo em cada ambiente, pode-se obter as somas de quadrados, ajustadas para  $\mu$ , para cada fonte de variação:

$$SQ_G = e \sum_{i=1}^g (\hat{g}_i)^2; \text{ com } (g - 1) \text{ graus de liberdade};$$

$$SQ_E = g \sum_{j=1}^e (\hat{e}_j)^2; \text{ com } (e - 1) \text{ graus de liberdade};$$

$$SQ_{Res} = \sum_{i=1}^g \sum_{j=1}^e \hat{\epsilon}_{ij}^2 \Leftrightarrow SQ_{G \times E} = \sum_{i=1}^g \sum_{j=1}^e (\widehat{ge})_{ij}^2 \text{ com } (g - 1)(e - 1) \text{ graus de liberdade};$$

Segundo Gauch e Zobel (1988) uma soma de quadrados da interação ( $SQ_{G \times E}$ ) pode ser considerada de elevada magnitude se esta representar pelo menos 20% da Soma de quadrados de tratamentos. Ressalta-se que, embora haja formas variadas de realizar o desdobramento da  $SQ_{G \times E}$  o objetivo é sempre de encontrar algum padrão que, ao ser incorporado ao modelo, possa melhorar a sua capacidade preditiva.

### 2.6.2. Desdobramento da soma de quadrado da interação por DVS

Na análise AMMI a decomposição  $SQ_{G \times E}$  pode ser particionada por DVS da seguinte forma (7):

$$\underbrace{\mathbf{GE}}_X = \underbrace{\mathbf{U}}_{P_{XX^T}} \underbrace{\mathbf{S}}_{\Lambda^{1/2}} \underbrace{\mathbf{V}^T}_{P_{X^T X}} = \sum_{k=1}^p \lambda_k \boldsymbol{\gamma}_k \boldsymbol{\alpha}_k \quad (7)$$

Com  $p = \min(g - 1), (e - 1)$ , em que  $\boldsymbol{\gamma}_k$  é o  $k$ -ésimo autovetor de  $(\mathbf{GE})(\mathbf{GE})^T$  e  $\boldsymbol{\alpha}_k$  é o  $k$ -ésimo autovetor de  $(\mathbf{GE})^T(\mathbf{GE})$ . De acordo com Duarte e Vencovsky (1999) a DVS pode ainda ser escrita como uma soma de  $p$  matrizes de posto unitário (8):

$$\underbrace{\mathbf{GE}}_{\begin{bmatrix} X=(x_{ij}) \\ (g \times e) \end{bmatrix}} = \underbrace{\lambda_1 \boldsymbol{\gamma}_1 \boldsymbol{\alpha}_1}_{\begin{bmatrix} X1=(x1_{ij}) \\ (g \times e) \end{bmatrix}} + \underbrace{\lambda_2 \boldsymbol{\gamma}_2 \boldsymbol{\alpha}_2}_{\begin{bmatrix} X2=(x2_{ij}) \\ (g \times e) \end{bmatrix}} + \dots + \underbrace{\lambda_p \boldsymbol{\gamma}_p \boldsymbol{\alpha}_p}_{\begin{bmatrix} Xp=(xp_{ij}) \\ (g \times e) \end{bmatrix}} \quad (8)$$

$$\sum_{i,j} X_{ij} = \sum_{i,j} (\widehat{ge})_{ij}^2 = \sum_{i,j} x1_{ij}^2 + \sum_{i,j} x2_{ij}^2 + \dots + \sum_{i,j} xp_{ij}^2 = SQ_{G \times E}$$

$$\sum_k \lambda_k^2 = \sum_{i,j} (\widehat{ge})_{ij}^2 = \lambda_1^2 + \lambda_2^2 + \dots + \lambda_p^2 = \sum_k SQ_{IPCK} = SQ_{G \times E}$$

Por  $\mathbf{GE}$  ser uma matriz de desvios centrados (possui média zero), a soma de quadrados de seus termos equivale à soma de quadrados da interação ( $SQ_{G \times E}$ ). Neste sentido, a soma de quadrados é desdobrada em componentes relativos a cada termo da DVS ou a cada eixo da interação da ACP que, conforme padrão internacional, pode ser denotada por IPCK. Pode-se



observar acima que não é estritamente necessário encontrar as  $p$  matrizes parciais, mas somente os autovalores de  $\mathbf{GE}$  ou, de forma equivalente, os autovalores de  $(\mathbf{GE})(\mathbf{GE})^T$  ou de  $(\mathbf{GE})^T(\mathbf{GE})$ . Para se encontrar a soma de quadrados relativa ao  $k$ -ésimo termo, ou eixo de interação, basta elevar ao quadrado o  $k$ -ésimo autovalor correspondente ( $\lambda_k^2$ ). O somatório de todos os  $\lambda_k^2$  recompõe a  $SQ_{G \times E}$ . Deve ser considerado que, dadas as propriedades para este tipo de decomposição, conclui-se que as  $p$  parcelas obtidas para a  $SQ_{G \times E}$  são ortogonais e independentes (DUARTE; VENCOVSKY, 1999)

Contudo o objetivo da análise AMMI não é necessariamente recompor toda a soma de quadrados da interação, mas somente a sua parte mais representativa. Portanto, pode ser que os dois ou três primeiros termos já sejam suficientes para explicar uma alta proporção da interação. Se isto for alcançado, é possível obter um modelo mais parcimonioso, consumindo um menor número de graus de liberdade, mas mantendo uma boa capacidade preditiva com relação a variabilidade devida a interação original. Assim sendo, poderá separar-se o comportamento padrão dos dados, ou seja, a possível lei que governa a interação, do ruído, isto é, o comportamento existente na interação que não possui interesse agrônômico e que pode comprometer a capacidade preditiva do modelo. Essa separação de padrão e ruído pode ser expressa da seguinte maneira (9):

$$\begin{aligned} SQ_{G \times E} &= \sum_{k=1}^p \lambda_k^2 = \left( \sum_{k=1}^n \lambda_k^2 \right) + \left( \sum_{k=n+1}^p \lambda_k^2 \right) \\ &= (SQ_{G \times E}[\text{padrão}]) + (SQ_{G \times E}[\text{ruído}]) \end{aligned} \quad (9)$$

A partir do número de componentes ou eixos singulares retidos no modelo, o modelo poderá ser denotado por diferentes formas:  $AMMI0, AMMI1, \dots, AMMIp$ . Em que:  $AMMI0$  representa apenas a parte aditiva do modelo e  $AMMIp$  contempla todos os possíveis componentes que poderiam ser retidos, ou seja, é o modelo completo de médias de caselas ( $n = p$ ). Este último seria o menos interessante em função de consumir muitos graus de liberdade e não resultar em otimização da informação que poderia ser extraída da interação.

### 2.6.3. Escolha do número de termos multiplicativos

Existem diferentes métodos que podem ser utilizados na escolha do número de termos a serem retidos no modelo. De acordo com Duarte e Vencovsky (1999), de forma geral, os

métodos costumam levar consideração, mesmo que indiretamente, a proporção da  $SQ_{G \times E}$  acumulada até o  $n$ -ésimo eixo ( $\sum_{k=1}^n \lambda_k^2 / SQ_{g \times e}$ ). Dias et al. (2013) argumentam que existem dois tipos de procedimentos mais utilizados: testes de significância dos termos multiplicativos e validação cruzada. Uma maneira muito utilizada de verificar a significância dos termos multiplicativos se dá pela atribuição de graus de liberdade à cada eixo da interação, de modo a tornar possível a obtenção do quadrado médio correspondente a cada eixo do modelo ( $QM_{IPCA_k}$ ) e, conseqüentemente, da estatística F, de maneira usual assim como é feita em uma análise de variância tradicional, a partir da razão  $QM_{IPCA_k} / QM_{erro\ médio}$ . A metodologia mais utilizada na atribuição de graus de liberdade a cada termo de modelo é o sistema proposto por Gollob (1968). Contudo, a validação cruzada também pode ser usada na escolha do número de termos multiplicativos.

### 2.6.3.1. Teste de Gollob

O teste proposto por Gollob (1968) é um dos métodos mais populares e mais utilizados devido a sua simplicidade e fácil aplicação, dispensando o uso de exaustivas simulações, como é o caso de outras metodologias como a proposta por Mandel (1971).

O teste consiste em atribuir graus de liberdade à soma de quadrados sequencial para o  $k$ -ésimo componente dos modelos AMMI:  $SQ_k = r\lambda_k^2$ , com  $k = 1, 2, \dots, p$ , em que  $p$  é o posto de  $(GE)$  e  $r$  o número de repetições. Contudo, a validade da distribuição F é questionável em função de os autovalores  $\lambda_k^2$  da matriz  $(GE)(GE)^T$  ou  $(GE)^T(GE)$  serem distribuídos como autovalores de uma matriz Wishart mas não possuem distribuição qui-quadrado, somado a questão de o teste assumir que  $r\lambda_k^2/\sigma^2$  também tem distribuição  $\chi^2$  (DIAS; KRZANOWSKI, 2003; DIAS, 2005; HONGYU et al., 2014).

O sistema de Gollob atribui graus de liberdade utilizando a definição estrutural clássica, isto é, o número de parâmetros estimados menos o número de restrições (GAUCH, 1992). De acordo com Duarte e Vencovsky (1999), ao se considerar o modelo AMMI, tem-se que dada eixo da interação ( $IPCA_k$ ) possui um autovalor  $\lambda_k$ ,  $g$  elementos para o vetor  $\boldsymbol{\gamma}_k$  e  $a$  elementos para o vetor  $\boldsymbol{\alpha}_k$ , resultando em  $g + a + 1$  parâmetros estimados. Retirando-se os efeitos principais, os autovetores  $\boldsymbol{\gamma}_k$  e  $\boldsymbol{\alpha}_k$  ficam sujeitos as restrições de  $\sum_{i=1}^g \gamma_{ik} = 0$  e  $\sum_{j=1}^a \alpha_{jk} = 0$ . Escalonando-os como vetores de comprimento unitários, outras duas restrições são impostas:  $\sum_{i=1}^g \gamma_{ik}^2 = 1$  e  $\sum_{j=1}^a \alpha_{jk}^2 = 1$ . Por fim, requerendo-se que cada eixo ou componente  $IPCA_n$ , com  $n > 1$ , seja

ortogonal a todos os  $n - 1$  eixos anteriores, são impostas  $2(n - 1)$  restrições:  $\sum_{i=1}^g \gamma_{in}\gamma_{ik} = 0$  e  $\sum_{j=1}^e \alpha_{jn}\alpha_{jk} = 0$ , com  $k = 1, 2, \dots, n - 1$ . Assim o eixo  $IPCA_n$  terá (10):

$$GL_{IPCA_n} = g + a + 1 - [2 + 2 + 2(n - 1)] = g + a - 1 - 2n \quad (10)$$

Portanto, o esquema da análise de variância segundo o sistema de Gollob considerando-se um delineamento inteiramente aleatorizado pode ser representado conforme a Tabela 3:

**Tabela 3.** Esquema da análise de variância considerando-se um delineamento inteiramente aleatorizado segundo o sistema proposto por Gollob

Fonte de Variação	Graus de liberdade	Soma de quadrados
Genótipos ( $g$ )	$g - 1$	$SQ_g$
Ambientes ( $e$ )	$e - 1$	$SQ_e$
Interação ( $ge$ )	$(g - 1)(e - 1)$	$SQ_{g \times e}$
IPC1	$g + e - 1 - (2 \times 1)$	$\gamma_1^2$
IPC2	$g + e - 1 - (2 \times 2)$	$\gamma_2^2$
IPC3	$g + e - 1 - (2 \times 3)$	$\gamma_3^2$
...		
IPC $p$	$g + e - 1 - (2 \times p)$	$\gamma_p^2$
Erro Médio/r	$ge(r - 1)$	$SQ_{\text{erro médio}}$
Total	$ger - 1$	$SQ_{\text{total}}$

### 2.6.3.2. Validação cruzada

A validação cruzada é um método livre de distribuição teórica que tem sido largamente utilizado em diferentes situações. Basicamente, consiste em dividir em dois subconjuntos os dados, contendo as repetições, para cada combinação entre genótipos e ambientes, seguindo algum critério aleatório: *i*) dados para ajuste do modelo AMMI; *ii*) dados para validação. As respostas do modelo AMMI são então comparadas com os dados de validação e a diferença entre os valores é calculada. Em seguida, calcula-se a raiz quadrada da soma de quadrados da diferença, dividida pelo número de valores das respostas preditas, denominada de diferença preditiva média ou RMSPD (11) (DUARTE; VENCOSKY, 1999; PIOVESAN; ARAÚJO; DIAS, 2009):

$$RMSPD = \left[ \left( \sum_{i=1}^n (\hat{y}_{ij} - y_{ij})^2 \right) / n \right]^{0,5} \quad (11)$$

Na validação cruzada, busca-se um pequeno valor de RMSPD, pois é um indicativo de um bom sucesso preditivo do modelo. Assim, o modelo escolhido será então selecionado definitivamente para analisar os dados completos com todas as repetições. Considerando-se essa

capacidade do método em conseguir prever dados de validação que não foram adicionados ao modelo, Gauch (1988) chama esta metodologia de avaliação preditiva, capaz de separar padrão e ruído, podendo resultar em uma precisão do modelo maior do que se fosse utilizado os dados completos com todas as repetições.

## 2.7. Limitações do modelo AMMI

O não atendimento às pressuposições da ANOVA implica na impossibilidade de aplicar a análise, não sendo possível, portanto, saber se a interação entre os fatores é significativa, assim como, inviabiliza escolher adequadamente o número de termos multiplicativos significativos, fazendo-se necessário utilizar técnicas adequadas para contornar esta limitação da análise. Métodos baseados em reamostragem são uma alternativa para escolher o número de termos significativos, além de a ANOVA e a DVS poderem ser substituídas por alternativas robustas à situações em que os pressupostos da análise de variância não são atendidos.

Um algoritmo baseado em reamostragem proposto por Malik, Forkman e Piepho (2019) é uma excelente alternativa para contornar as limitações do modelo AMMI tradicional, em função de usar *bootstrap* não paramétrico e permutação. Sendo assim, não é necessário o atendimento de nenhuma pressuposição forte como normalidade e homoscedasticidade da variância residual. Os autores propõem um teste robusto à presença de heterogeneidade de variâncias e o comparam com outros testes paramétricos, sendo este superior ao *bootstrap* paramétrico quando as pressuposições da ANOVA são violadas.

Uma outra abordagem que pode ser empregada é a aplicação de um modelo AMMI robusto, denominado de R-AMMI, proposto por Rodrigues, Monteiro e Lourenço (2016). O modelo é uma generalização da abordagem clássica, onde métodos estatísticos robustos são utilizados para substituir a ANOVA e a DVS, com o objetivo de modelar, estruturar e analisar a interação GE. No estudo, os autores verificam a robustez do modelo diante de vários cenários de contaminação que podem estar presente no conjunto de dados, por meio de um estudo de simulação de Monte Carlo.

### 2.7.1. Método baseado em reamostragem (resampling-based)

Este método foi proposto por Malik, Forkman e Piepho (2019) para dados replicados. Uma abordagem específica para dados não replicados pode ser consultada em Malik et al. (2018). A estatística do teste para os  $K + 1$  termos é (12):

$$T_r = \hat{\lambda}_{K+1}^2 / s^2 \quad (12)$$

Em que  $\hat{\lambda}_{K+1}^2$  é a estimativa do autovalor  $\lambda_{K+1}^2$ , obtido a partir da decomposição em valor singular e  $s^2$  é proveniente da razão entre o quadrado médio do resíduo obtido a partir da ANOVA e o número de repetições ou blocos, representando o quadrado médio de erro agrupado com base na média genótipo-ambiente.

A partir de então, podem ser simuladas amostras, sob hipótese nula, utilizando algum método de reamostragem como o *bootstrap* não paramétrico ou permutação. Inicialmente, ajusta-se o modelo completo com genótipos, ambientes e blocos, se for o caso, considerando-se todos de efeitos fixos. Os resíduos deste modelo  $\varepsilon_{ijr}$ , armazenados em uma matriz  $\mathbf{R}$ , são computados em uma base replicada. A partir de então, um grande número,  $B$ , de amostras é extraído do modelo ajustado sob a hipótese nula. Em todas as amostras, a soma,  $\hat{\Theta}_{(k)}$ , dos  $K$  primeiros termos multiplicativos da DVS são considerados a verdadeira interação fixa. O erro residual da  $b$ -ésima amostra ( $\hat{\mathbf{R}}_b^B$ ) é sorteado aleatoriamente de  $\mathbf{R}$  e adicionado ao efeito da interação  $\hat{\Theta}_{(k)}$  para obtenção dos dados simulados. Se os resíduos forem extraídos com reposição, então, tem-se a  $b$ -ésima amostra *bootstrap* não paramétrica, mas se forem extraídos sem reposição, será obtida a  $b$ -ésima amostra baseada em permutação. A estatística do teste  $T_r$  é computada para a  $b$ -ésima amostra para obter a  $b$ -ésima estatística do teste simulada  $T_{rb}$ , que então é comparada com a estatística  $T_r$  calculada a partir dos dados originais. Por fim, é calculado o  $p$ -valor que é obtido empiricamente a partir da proporção de estatísticas  $T_{rb}$  ( $b = 1, \dots, B$ ) maior que  $T_r$ .

O algoritmo baseado em reamostragem pode ser descrito da seguinte maneira:

- i) Ajustar o modelo contendo os efeitos fixos de ambientes, genótipos e blocos, no caso de um modelo AMMI, ou sem o efeito de genótipo no modelo GGE<sup>1</sup>. Em seguida, calcular  $s^2$  e obter os resíduos  $\varepsilon_{ijr}$  que devem ser armazenados em uma matriz  $\mathbf{R}$ .
- ii) Obter os dados simulados adicionando ( $\hat{\mathbf{R}}_b^B$ ) em  $\hat{\Theta}_{(k)}$ , sob hipótese nula.
- iii) Aplicar a DVS em  $\hat{\Theta}_{(k)}$ , obtida a partir dos dados simulados e calcular  $T_{rb}$  da mesma maneira que  $T_r$ .
- iv) Obter o valor-p empírico a partir da frequência relativa de  $T_{rb} > T_r$

---

<sup>1</sup> GGE é uma técnica multivariada semelhante à análise AMMI que também é capaz de trabalhar com dados provenientes de ensaios de genótipos avaliados em diferentes ambientes.

### 2.7.2. R-AMMI

Rodrigues, Monteiro e Lourenço (2016) propuseram um modelo AMMI robusto que é executado em duas etapas: i) Ajustar um modelo de regressão robusto usando o estimador M-huber (HUBER, 1981) para substituir o modelo ANOVA. ii) Adaptar a decomposição em valor singular usando diferentes métodos que consideram a ocorrência de observações discrepantes na matriz de interação. Este método, denominado inicialmente de AL1-SVD, foi proposto por Hawkins, Liu e Young (2001). O objetivo é conseguir trabalhar com matrizes retangulares que contenham uma minoria de células discrepantes ou dados ausentes, aplicando a norma L1 em vez da norma L2 de mínimos quadrados mais usual, para calcular uma aproximação robusta da DVS. O algoritmo funciona da seguinte forma:

- i) Começar com uma estimativa inicial do autovetor coluna principal  $\mathbf{a}_1$ ;
- ii) Para cada coluna  $j, j = 1, \dots, p$ , deve ser ajustado o coeficiente  $c_j$  da regressão L1 por  $\min \sum_{i=1}^n |x_{ij} - c_j a_{i1}|$ ;
- iii) Calcular a estimativa resultante do autovetor linha  $\mathbf{b}_1 = \mathbf{c}/\|\mathbf{c}\|$ , onde  $\|\cdot\|$  se refere a norma euclidiana;
- iv) Usando a estimativa do autovetor linha, a estimativa do autovetor coluna deve ser refinada. Para cada coluna  $i, i = 1, \dots, n$ , deve ser ajustado o coeficiente  $d_j$  da regressão L1 por  $\min \sum_{j=1}^p |x_{ij} - d_i b_{j1}|$ ;
- v) Calcular a estimativa resultante do autovetor coluna  $\mathbf{a}_1 = \mathbf{d}/\|\mathbf{d}\|$ ;
- vi) Retornar ao passo 2 e repetir o processo iterativamente até se alcançar convergência.

As regressões L1 envolvidas são triviais, reduzindo-se a encontrar medianas ponderadas. Que é resolvido pela mediana ponderada das razões  $x_{ij}/a_{1j}$  com pesos  $|a_{1j}|$ . Se  $a_{1j} = 0$ , então o termo pode ser ignorado, pois não contribui para o peso. As células ausentes são omitidas do cálculo da mediana (13).

$$\min \sum_{i=1}^n |x_{ij} - c_j a_{i1}| = \min \sum_{i=1}^n |a_{1j}| \cdot |x_{ij}/a_{1j} - c_j| \quad (13)$$



### 3. MATERIAL E MÉTODOS

#### 3.1. Área experimental

Os dados são provenientes de um programa de melhoramento genético de trigo realizado em uma estação experimental do CYMMIT (*Centro Internacional de Mejoramiento de Maíz y Trigo*), localizada em Ciudad Obregon, Sonora, no noroeste do México (lat. 27,33 N, long. 109,09 W, 38 masl). A região é caracterizada por um clima árido, com precipitação média anual de 308 mm, e temperatura média anual de 23,5°C. O solo é classificado como Vertissolo hipossódico (Calcarico, Crômico), considerando-se a base de referência mundial (VERHULST; DECKERS; GOVAERTS, 2009). Os ensaios são realizados anualmente no período de novembro a maio. O banco de dados é de acesso público, portanto, informações mais detalhadas podem ser consultadas diretamente em Honsdorf et al. (2022).

#### 3.2. Desenho experimental e material vegetal

Na área experimental há quatro ambientes agronômicos: (E1) canteiros convencionalmente arados com irrigação total, (E2) canteiros convencionalmente arados com irrigação reduzida, (E3) plantio direto com irrigação total e (E4) plantio direto com irrigação reduzida. Todos os ambientes foram irrigados no momento da semeadura. Os ambientes com irrigação total receberam três ou quatro irrigações adicionais (aproximadamente 520 mm de água por estação) com o intuito de evitar o estresse hídrico; os ambientes com irrigação reduzida receberam uma irrigação adicional (aproximadamente 240 mm por estação).

Dentro de cada ambiente foi implantado um delineamento aleatorizado em blocos, com três blocos e 24 genótipos de duas espécies de trigo, sendo 12 genótipos de trigo comum (*Triticum aestivum*) e 12 de trigo durum (*Triticum turgidum*). A variável resposta que será analisada neste estudo é o rendimento de grãos (kg/ha), referentes à safra de 2017-2018. Os genótipos analisados podem ser verificados na Tabela 4.



**Tabela 4.** Genótipos de trigo comum (*Triticum aestivum*) e trigo durum (*Triticum turgidum*).

Trigo durum		Trigo comum	
Código	Nome	Código	Nome
G1	Siete Cerros 66	G23	COCORIT C71
G2	Pavon 76	G24	MEXICALI C75
G3	Seri 82	G25	YAVAROS C79
G10	Nesser	G26	ALTAR C84
G4	Super Kauz	G28	ATIL C2000
G8	Attila	G29	JUPARE C2001
G5	Baviacora 92	G30	SAMAYOA C2004
G6	Cachanilla 2000	G41	SOOTY_9/...
G7	Tacupeto 2001	G47	CIRNO C2008
G11	Weebill#1	G67	TOPDY_18/...
G39	NAVAJOA 2007	G68	BAROYECA C2013
G16	Becard	G69	ARTICO/...

### 3.3. Análise estatística

Considerou-se três conjunto de dados, sendo um formado por ambas as espécies (24 genótipos) e outros dois subconjuntos contendo 12 genótipos cada um, referente à cada espécie estudada: trigo durum (*Triticum turgidum* L.) e trigo comum (*Triticum aestivum* L.).

Inicialmente foi realizada uma avaliação inicial (análise exploratória dos dados, resíduos, verificação das pressuposições por meio de testes estatísticos e transformação). Aplicou-se a análise exploratórias dos dados e resíduos para estudar a existência de interação entre os fatores, e verificação dos pressupostos da análise de variância. Na análise de resíduos, valores fora do intervalo compreendido entre -3 e 3 foram considerados como observações discrepantes. Para testar a normalidade dos dados, aplicou-se o teste de Shapiro e Wilk (1965). Em relação à homogeneidade de variâncias, aplicou-se o teste de Breusch e Pagan (1979) para os experimentos individuais e, na análise conjunta, aplicou-se os testes de Hartley (1950) e Bartlett (1937) para testar a homogeneidade de variâncias dos efeitos principais e da combinação genótipo × ambiente. Nos casos em que os pressupostos não foram atendidos, procedeu-se com a transformação utilizando-se o método de Box e Cox (1964).

Em seguida, realizou-se a ANOVA individual dos experimentos quando os ambientes atenderam aos pressupostos da análise. O modelo AMMI foi aplicado quando às pressuposições de normalidade e heterocedasticidade foram atendidos. Quando não foi possível atender aos pressupostos de normalidade de resíduos e homoscedasticidade de variância residual de cada efeito principal (genótipo, ambiente e combinação genótipo × ambiente), aplicou-se o algoritmo baseado em reamostragem (MALIK; FORKMAN; PIEPHO, 2019) para verificar o número de termos multiplicativos significativos, em seguida, construiu-se o *Biplot* usando DVS normalmente. Ainda nessa etapa, o algoritmo R-AMMI (RODRIGUES; MONTEIRO; LOURENÇO, 2016)

também foi usado para realizar a análise AMMI robusta, adotando-se o mesmo número de termos multiplicativos significativos sugerido por reamostragem.

### **3.4. Pacotes estatísticos**

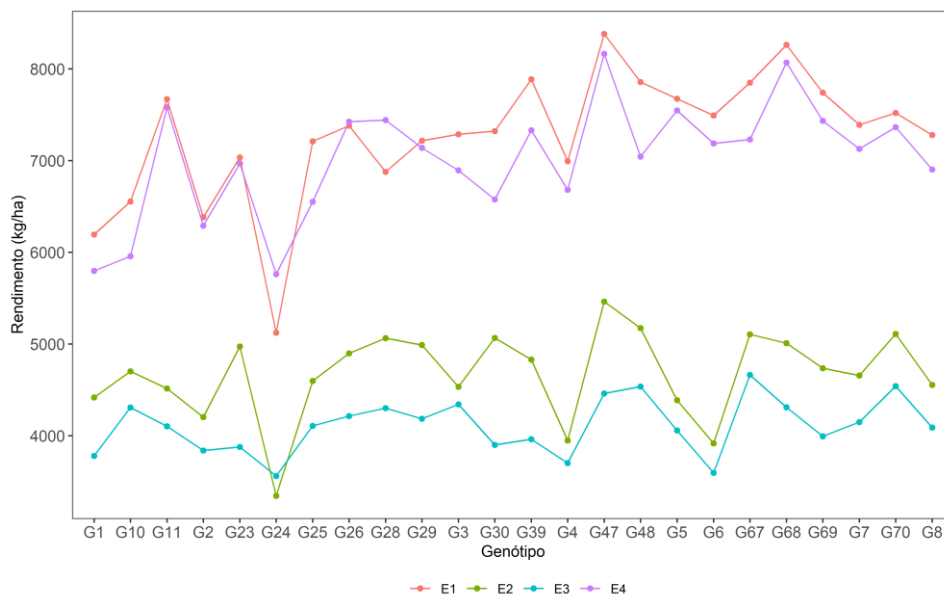
Todas as análises foram realizadas no software R (R CORE TEAM, 2021). A análise AMMI tradicional foi realizada usando-se o pacote *metan* (OLIVOTO; LÚCIO, 2020) com as funções *performis\_amm*i e *plot\_scores*. A análise AMMI robusta foi realizada usando-se o pacote *geneticae* (ANGELINI; PRUNELLO; CERVIGNI, 2022) com a função *rammi* indicando a tabela de médias de cada conjunto de dados.



## 4. RESULTADOS E DISCUSSÃO

### 4.1. Análise exploratória

O rendimento dos genótipos de arroz em cada ambiente pode ser observado na Figura 1. É possível perceber que os ambientes um e quatro são os mais produtivos, sendo visualmente similares, mas ambos contrastando com os ambientes dois e três.



**Figura 1.** Rendimento de grãos de trigo (kg/ha) por genótipos segundo os ambientes.

Analisando-se mais detalhadamente cada genótipo, percebe-se que todos possuem o mesmo comportamento nos ambientes, independentemente da espécie. Isto é, rendimentos maiores em E1 e E4, bem como, menores rendimentos em E2 e E3. Assim, pode-se perceber que os ambientes estão influenciando no desempenho produtivo dos genótipos (Figura 2).

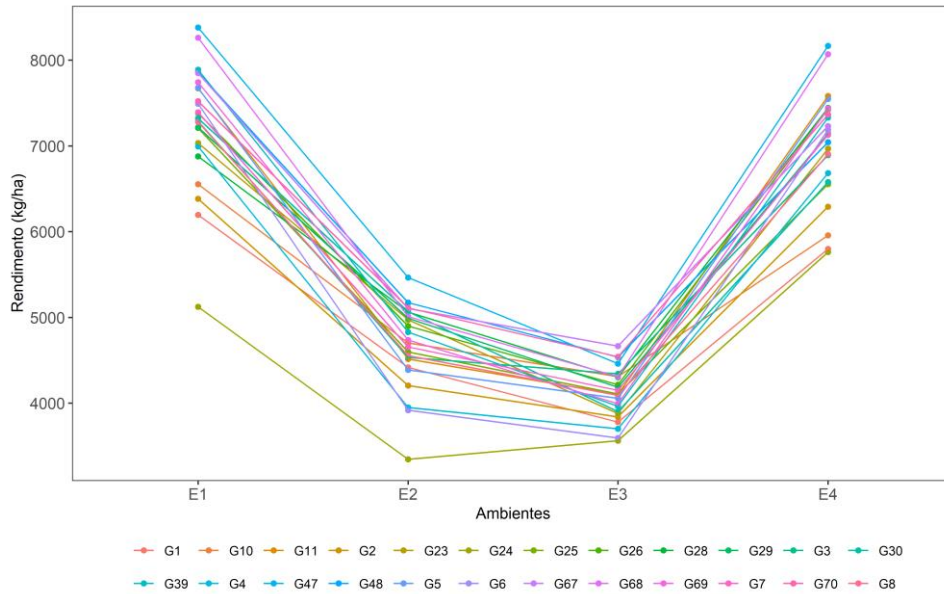
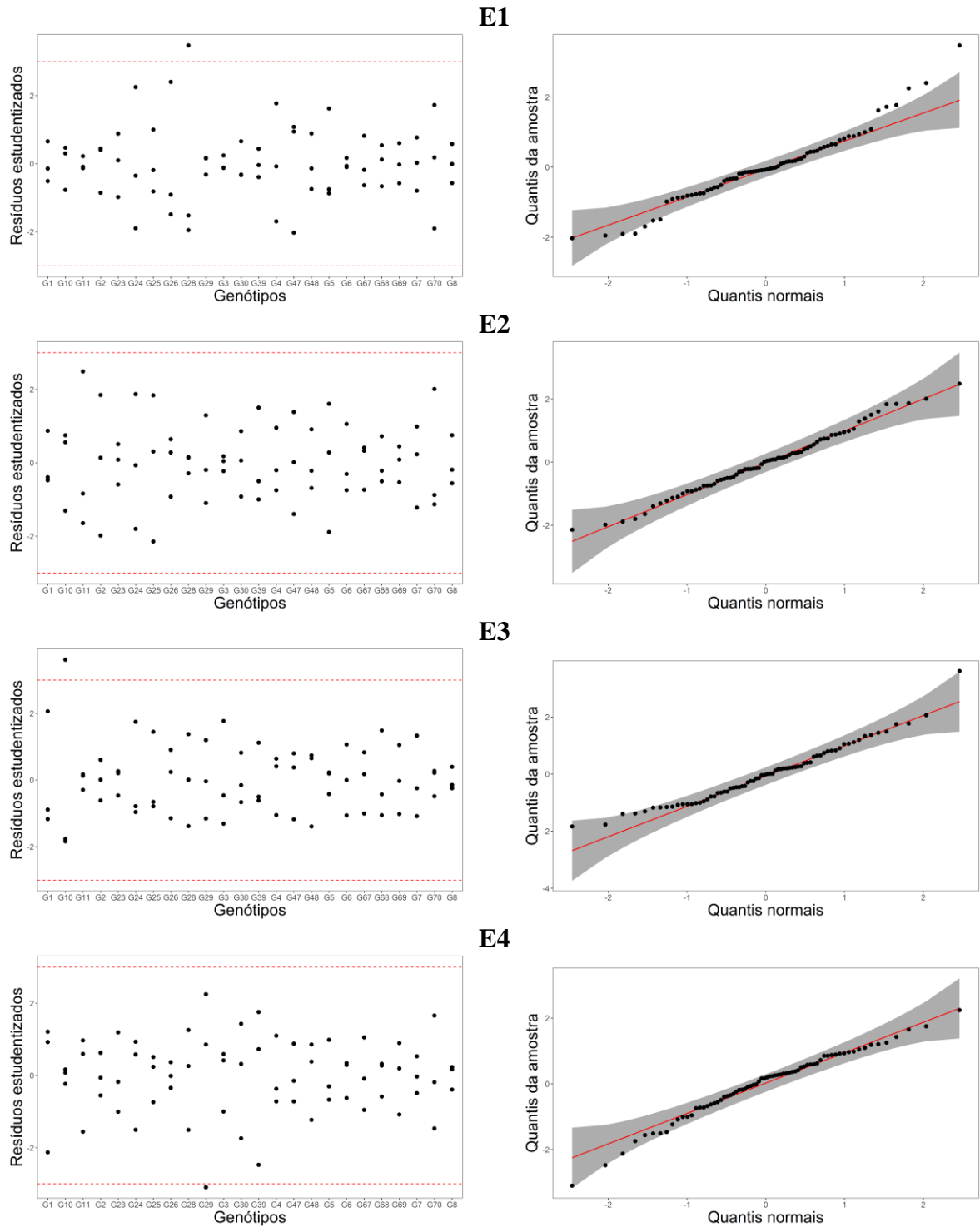


Figura 2. Rendimento de grãos de trigo (kg/ha) por ambientes segundo os genótipos.

## 4.2. Análise de resíduos

### 4.2.1. Dados completos

Fazendo-se uma exploração gráfica dos resíduos de acordo com cada ambiente, é possível perceber que os dados aparentemente possuem maiores problemas relacionados à homogeneidade de variâncias entre os genótipos do que em relação à normalidade. Com exceção do ambiente um, os demais ambientes não aparentam ter problemas graves com relação à distribuição dos dados. Em E1, um resíduo relacionado ao genótipo 28 pode ser considerado uma observação discrepante, em função de estar fora do intervalo compreendido entre -3 e 3, assim como ocorre com um resíduo relacionado ao G10, no ambiente três e o G29 no ambiente quatro (Figura 3).



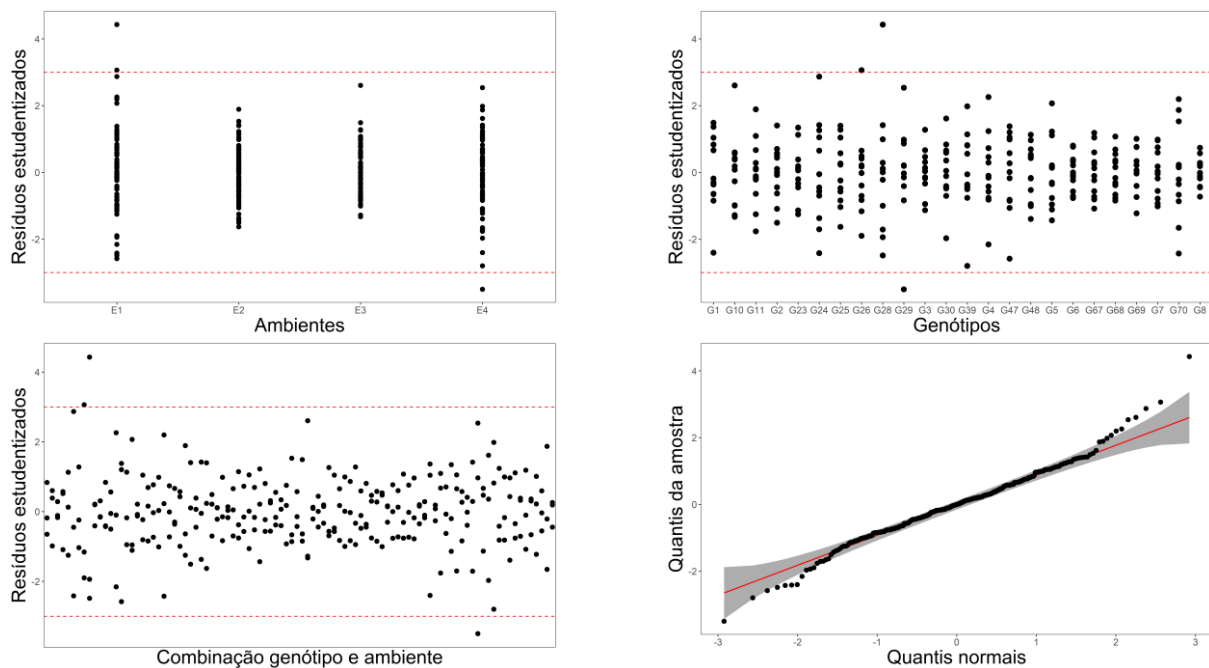
**Figura 3.** Gráficos de genótipos *versus* resíduos estudentizados (coluna à esquerda) e quantis normais *versus* quantis observados na amostra (coluna à direita) para cada ambiente, considerando-se os dados completos.

Aplicando-se os testes de Shapiro-Wilk e Breush-Pagan para verificar a normalidade e a homogeneidade de variâncias, respectivamente, percebe-se que somente no ambiente 2 há o atendimento de todas as pressuposições da ANOVA (Tabela 5).

**Tabela 5.** Testes de normalidade e homogeneidade de variâncias para a variável rendimento de grãos (kg/ha) para os dados completos.

Locais	Shapiro-Wilk		Breusch-Pagan	
	Estatística ( <i>W</i> )	Valor-p	Estatística ( <i>BP</i> )	Valor-p
1	0,959	0,019	42,286	0,017
2	0,991	0,889	36,507	0,064
3	0,966	0,045	44,233	0,010
4	0,977	0,217	38,97	0,037

Analisando-se os dados conjuntamente, verifica-se que os pressupostos de normalidade e homogeneidade de variâncias, seja para ambiente, genótipos ou a combinação genótipo ambiente também não são atendidos. Isto pode ser observado tanto graficamente (Figura 4), quanto por meio de teste estatístico (Tabela 6). A única exceção é a existência de homogeneidade de variâncias apontada pelo teste de Hartley para a combinação genótipo e ambiente.

**Figura 4.** Gráficos de resíduos dos dados completos para ambientes, genótipos e combinação genótipo e ambiente.**Tabela 6.** Testes de homogeneidade de variâncias para cada fonte de variação e normalidade dos resíduos.

Fontes de Variação	Bartlett		Hartley		Shapiro-Wilk	
	$\chi^2$	Valor-p	<i>Fmax</i>	<i>p</i> -valor	<i>W</i>	Valor-p
Ambiente (E)	130,72	0,008	3,108	0,000		
Genótipo (G)	33,089	0,000	18,047	0,005		
Gen:Amb (GE)	56,624	0,000	591,56	0,557		
Dados completos					0,980	0,000

Procurou-se estudar a possibilidade de transformação dos dados. No gráfico de valores esperados médios (preditos) *versus* resíduos estudentizados é possível perceber que o padrão assumido pode não ser passível de transformação por Box-Cox (Figura 5).

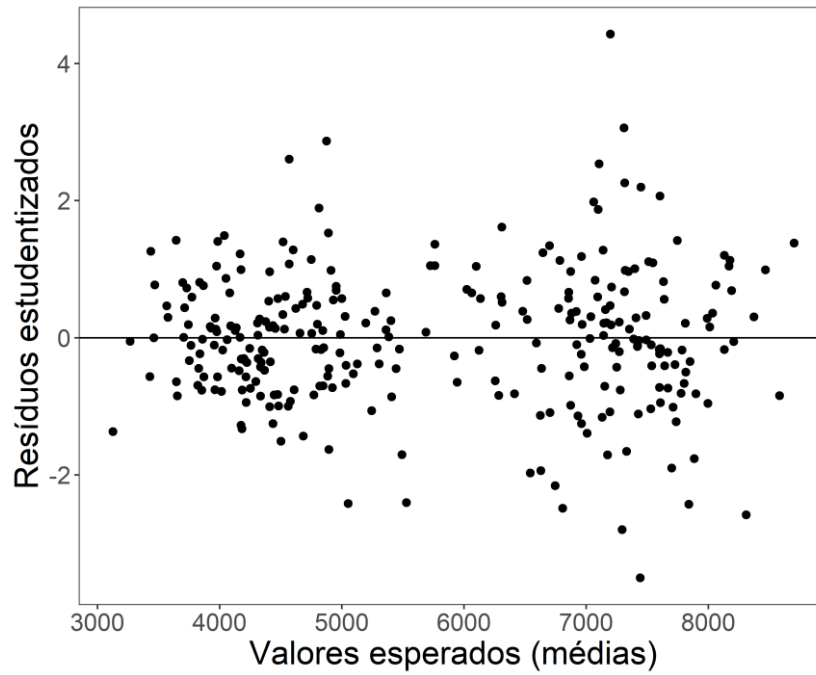


Figura 5. Resíduos estudentizados *versus* valores esperados.

Contudo, procedeu-se com a transformação de Box-Cox elevando o rendimento a 0,3 ( $y^{0,3}$ ), conforme sugerido. Observou-se que ocorreu uma leve melhora em relação aos genótipos e na distribuição dos dados (Figura 6)

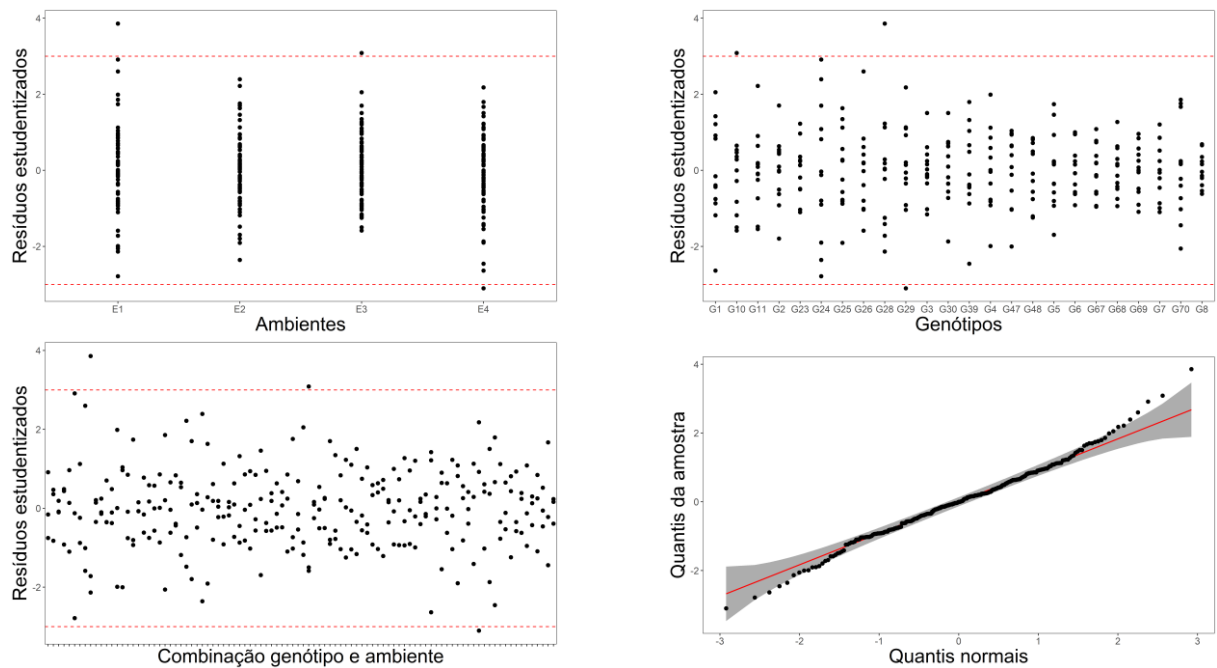


Figura 6. Gráficos de resíduos dos dados completos após transformação para ambientes, genótipos e combinação genótipo e ambiente.



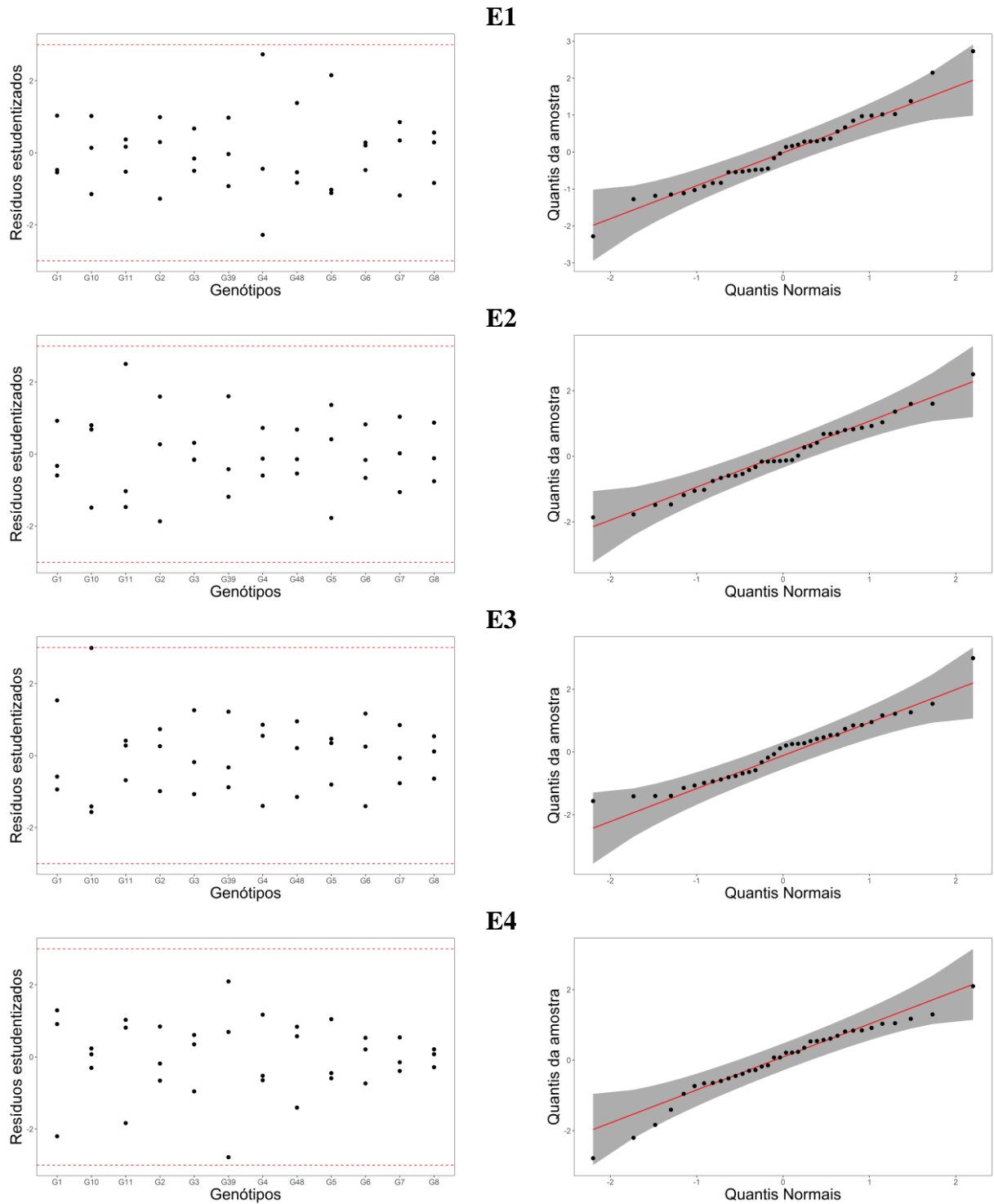
Com a aplicação dos testes estatísticos após a transformação dos dados, verificou-se que pelo teste de Bartlett as variâncias permanecem heterogêneas para cada fonte de variação. Contudo, pelo teste de Hartley, com exceção de genótipos, as demais fontes de variação passaram a ter variâncias homogêneas. No entanto, não foi possível atender ao pressuposto de normalidade pelo teste de Shapiro-Wilk (Tabela 7).

**Tabela 7.** Testes de homogeneidade de variâncias para cada fonte de variação e normalidade dos resíduos

Fontes de Variação	Bartlett		Hartley		Shapiro-Wilk	
	$\chi^2$	Valor-p	<i>Fmax</i>	Valor-p	<i>W</i>	Valor-p
Ambiente (E)	4,527	0,210	1,564	0,240		
Genótipo (G)	55,715	0,000	18,514	0,004		
Gen:Amb (GE)	123,6	0,026	655,83	0,5213		
Dados completos					0,989	0,025

#### 4.2.2. Trigo comum (*Triticum aestivum*)

Analisando-se os locais individualmente considerando apenas trigo comum, observa-se que os dados não apresentam grandes desvios relacionados à normalidade dos resíduos, assim como, com relação a observações discrepantes (Figura 7).



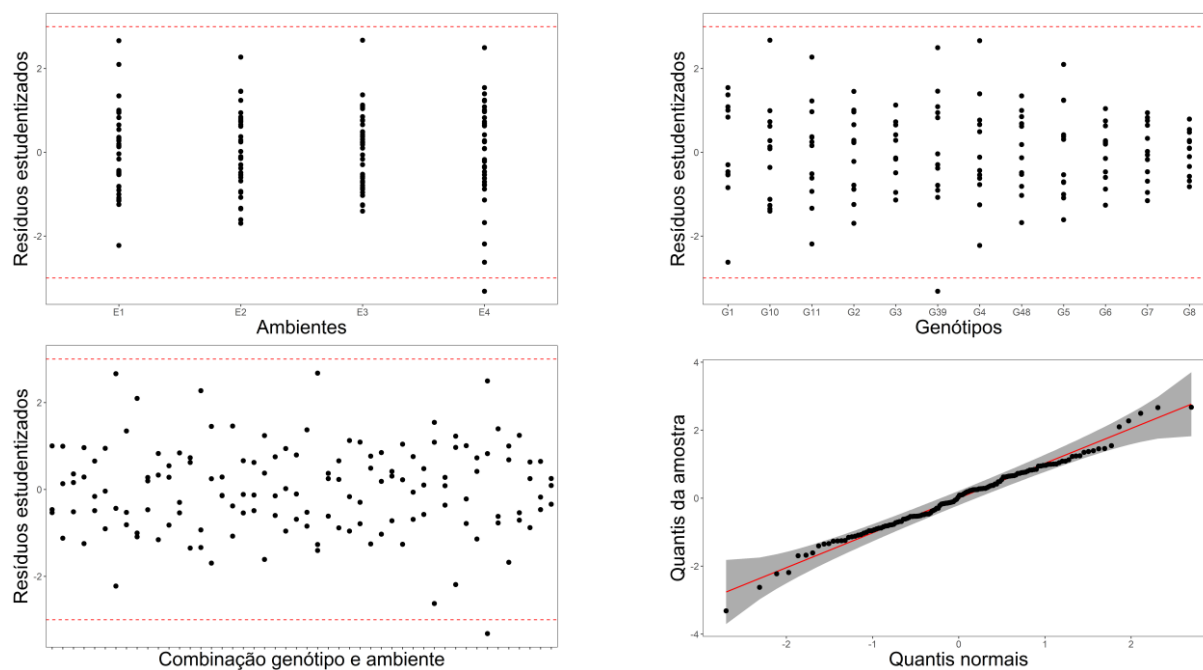
**Figura 7.** Gráficos de genótipos *versus* resíduos estudentizados (coluna à esquerda) e quantis normais *versus* quantis observados na amostra (coluna à direita) para cada ambiente, considerando-se os dados de trigo comum (*Triticum aestivum*).

Com a aplicação dos testes estatísticos para estudar a normalidade e a homogeneidade de variâncias dos resíduos, observa-se que todos os ambientes atendem aos pressupostos da análise de variância, com exceção do ambiente três que não apresentou homogeneidade de variâncias (Tabela 8).

**Tabela 8.** Testes de normalidade e homogeneidade de variâncias para a variável rendimento de grãos (kg/ha) para os dados de trigo comum (*Triticum aestivum*).

Locais	Shapiro-Wilk		Breusch-Pagan	
	Estatística (W)	Valor-p	Estatística (BP)	Valor-p
E1	0,972	0,490	20,471	0,084
E2	0,984	0,863	18,416	0,142
E3	0,955	0,149	23,398	0,037
E4	0,960	0,216	21,269	0,068

Analisando-se conjuntamente os dados de *T. aestivum* também não se observa graves desvios graficamente. No entanto, o resíduo do genótipo 39 pode ser classificado como uma observação discrepante. Quando se analisa a combinação genótipo e ambiente também é possível perceber que há uma observação discrepante (Figura 8).



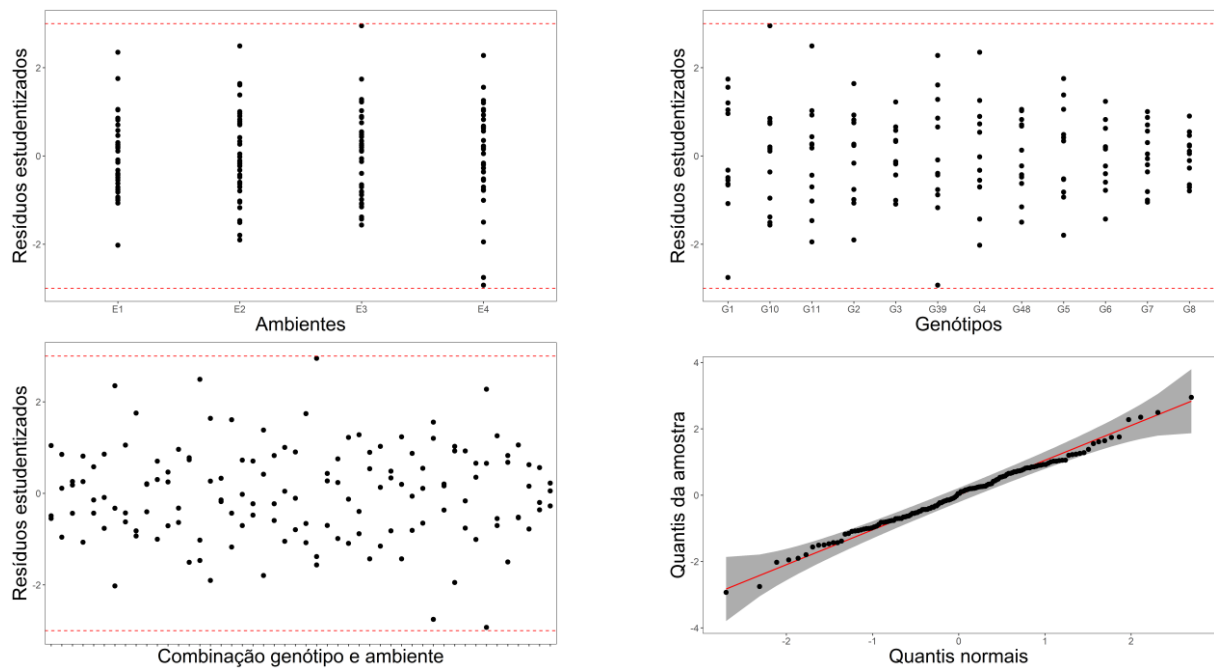
**Figura 8.** Gráficos de resíduos relacionados aos dados de *T. aestivum* para ambientes, genótipos e combinação genótipo e ambiente.

Aplicando-se os testes estatísticos, tem-se que os resíduos seguem distribuição normal. Mas em relação a homogeneidade de variâncias, somente os genótipos apresentam variâncias heterogêneas (0).

**Tabela 9.** Testes de homogeneidade de variâncias para cada fonte de variação e normalidade dos resíduos relacionados a espécie *T. aestivum*.

Fontes de Variação	Bartlett		Hartley		Shapiro-Wilk	
	$\chi^2$	Valor-p	<i>Fmax</i>	Valor-p	<i>W</i>	Valor-p
Ambiente (E)	3,798	0,284	1,767	0,342		
Genótipo (G)	22,033	0,024	8,568	0,043		
Gen:Amb (GE)	44,914	0,559	147,32	0,748		
Dados completos					0,988	0,265

Aplicou-se em seguida a transformação por Box-Cox, elevando-se o rendimento a 0,462 ( $y^{0,062}$ ). Após a transformação dos dados, verifica-se que não há mais a ocorrência de resíduos discrepantes (Figura 9).



**Figura 9.** Gráficos de resíduos relacionados aos dados transformados de trigo comum (*T. aestivum*) para ambientes, genótipos e combinação genótipo e ambiente.

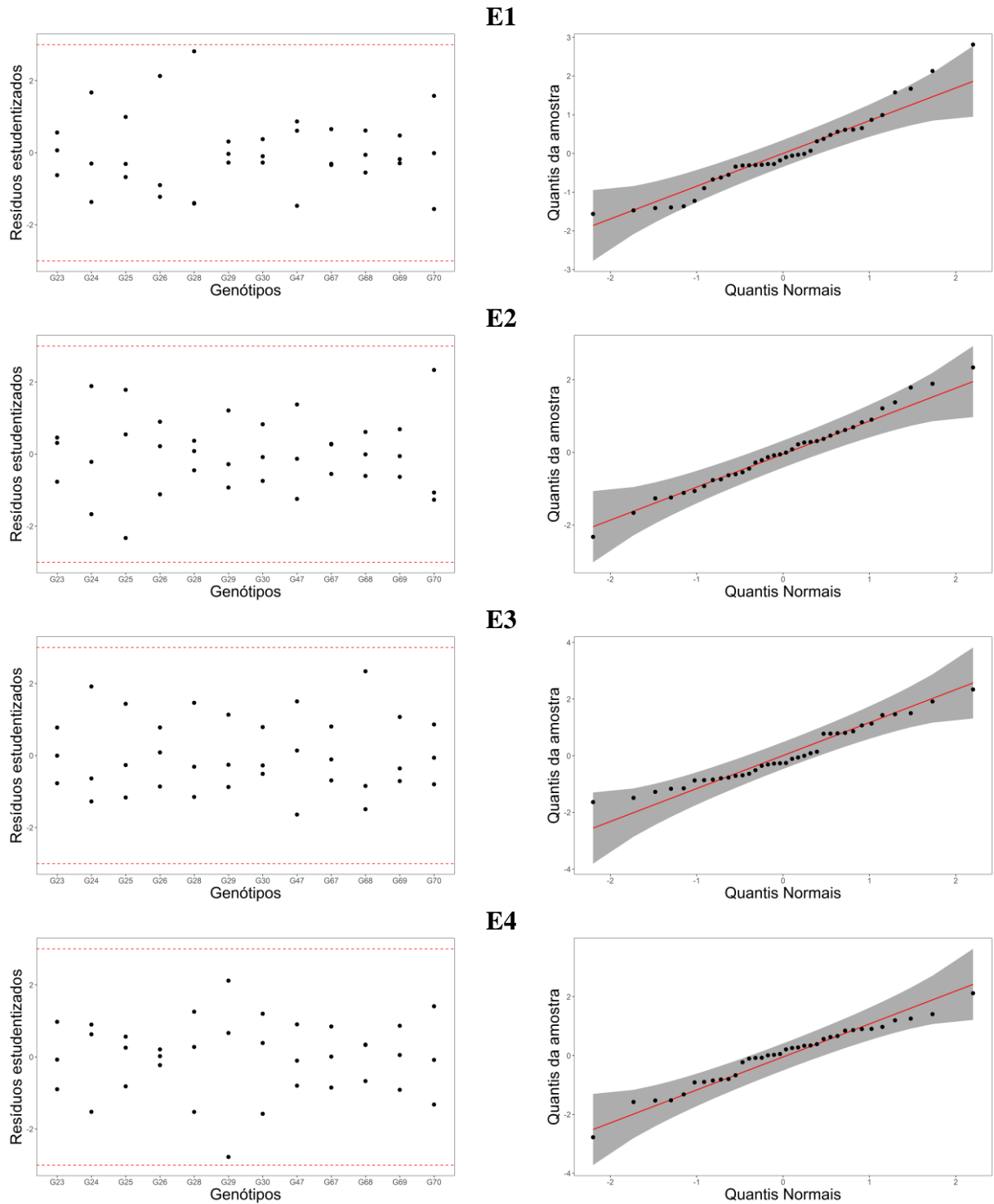
Aplicando-se os testes estatísticos sobre os dados transformados, percebeu-se que houve discordância entre os testes de Hartley e Bartlett com relação a homogeneidade de variâncias entre os genótipos. Os resíduos permanecem seguindo distribuição normal (Tabela 10).

**Tabela 10.** Testes de homogeneidade de variâncias para a variável rendimento de grãos (kg/ha) transformada, para cada fonte de variação e normalidade dos resíduos para a espécie *T. aestivum*.

Fontes de Variação	Bartlett		Hartley		Shapiro-Wilk	
	$\chi^2$	Valor-p	<i>Fmax</i>	Valor-p	<i>W</i>	Valor-p
Ambiente (E)	2,391	0,500	1,682	0,424		
Genótipo (G)	29,915	0,034	7,165	0,089		
Gen:Amb (GE)	44,624	0,572	110,73	0,838		
Dados completos					0,994	0,774

#### **4.2.3. Trigo durum (*Triticum turgidum*)**

Fazendo-se uma exploração gráfica dos resíduos, observa-se que não há a ocorrência de pontos discrepantes ou problemas relacionados a normalidade dos dados nos ambientes (Figura 10)



**Figura 10.** Gráficos de genótipos *versus* resíduos estudentizados (coluna à esquerda) e quantis normais *versus* quantis observados na amostra (coluna à direita) para cada ambiente, considerando-se os dados de trigo durum (*Triticum turgidum*).

Aplicando-se os testes estatísticos para verificar o atendimento aos pressupostos da análise de variância, verifica-se que as pressuposições foram atendidas (Tabela 11). Assim, prosseguiu-se com a análise de variância individual por locais.

**Tabela 11.** Testes de normalidade e homogeneidade de variâncias para a variável rendimento de grãos (kg/ha), considerando-se os dados de trigo durum (*T. turgidum*).

LOCAIS	SHAPIRO-WILK		BREUSCH-PAGAN	
	Estatística ( <i>W</i> )	Valor-p	Estatística ( <i>BP</i> )	Valor-p
E1	0,949	0,095	20,612	0,081
E2	0,992	0,995	18,75	0,131
E3	0,953	0,134	17,653	0,171
E4	0,972	0,488	18,517	0,139

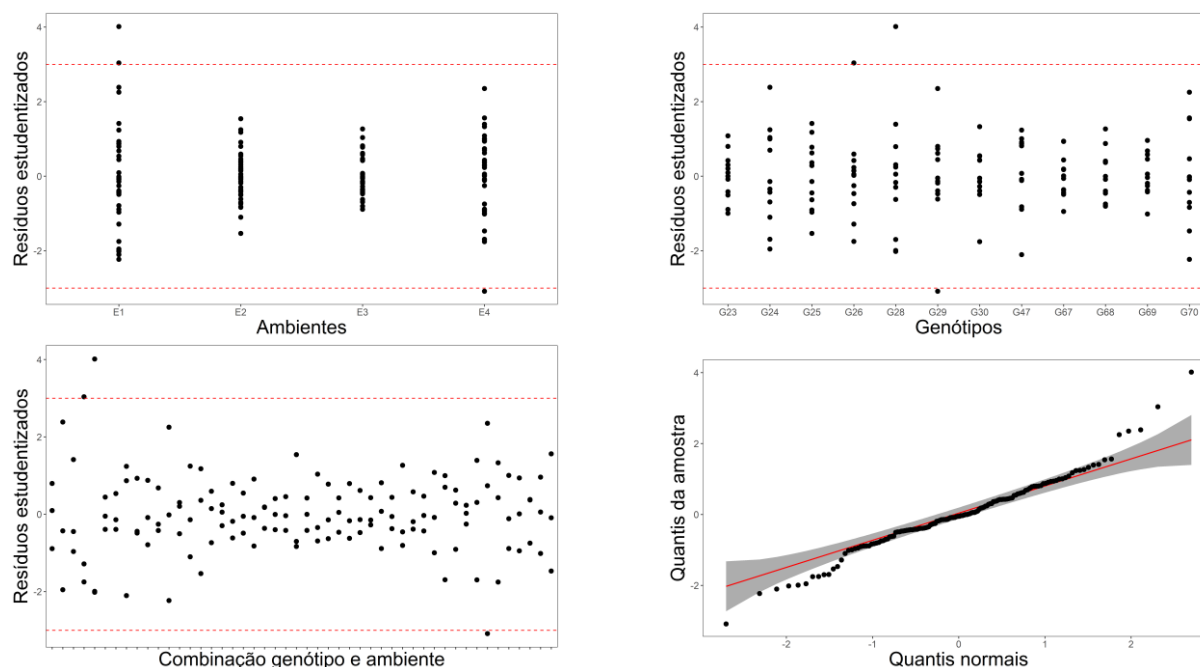
De acordo com a análise de variância, todos os experimentos apresentaram diferenças significativas entre genótipos, variando os níveis de significância de 5% a 0,1% (Tabela 12).

**Tabela 12.** Resumo da análise de variância para a variável rendimento de grãos (Kg/ha), por locais, considerando-se os genótipos de trigo durum (*T. turgidum*).

FV	GL	QM L1	QM L2	QM L3	QM L4
Blocos (B)	2	963721*	1005753***	290643**	778026*
Genótipos (G)	11	2083781***	818556***	291640***	1312182***
Resíduo	22	257569	55057	37099	156177
Total	35	3305071	1879366	619382	2246385

FV: Fonte de variação; GL: Graus de liberdade; QM: Quadrado médio. \* Teste F significativo a 5% de probabilidade; \*\* Teste F significativo a 1% de probabilidade; \*\*\* Teste F significativo a 0,1% de probabilidade.

Realizando-se a análise conjunta dos dados, é possível perceber que há a ocorrência de valores discrepantes e problemas relacionados à normalidade dos dados (Figura 11).



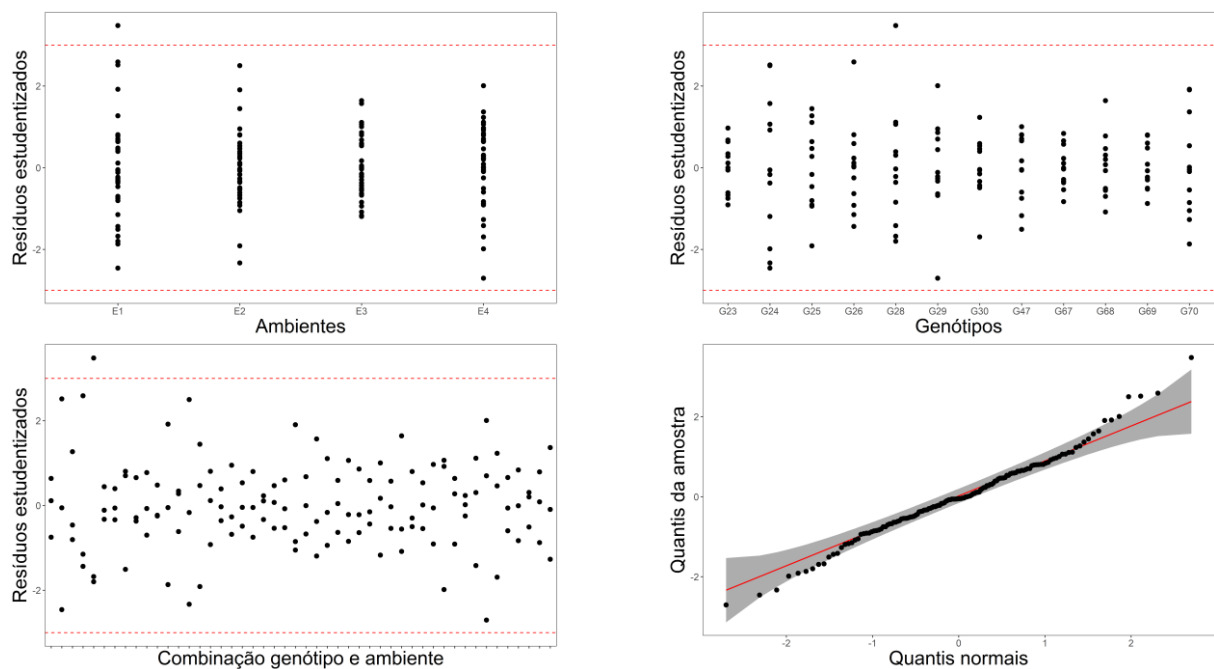
**Figura 11.** Gráficos de resíduos relacionados aos dados de trigo durum (*T. turgidum*) para ambientes, genótipos e combinação genótipo e ambiente.

O não atendimento às pressuposições da ANOVA é confirmado a partir da aplicação dos testes estatísticos, com exceção da existência de homogeneidade de variâncias para a combinação genótipo  $\times$  ambiente segundo o teste de Hartley (Tabela 13).

**Tabela 13.** Testes de homogeneidade de variâncias para a variável rendimento de grãos (kg/ha), para cada fonte de variação e normalidade dos resíduos, considerando-se os dados de trigo durum (*T. turgidum*).

Fontes de Variação	Bartlett		Hartley		Shapiro-Wilk	
	$\chi^2$	Valor-p	<i>F</i> max	Valor-p	<i>W</i>	Valor-p
Ambiente (E)	39,291	0,000	6,943	0,000		
Genótipo (G)	29,996	0,002	8,502	0,045		
Gen:Amb (GE)	66,053	0,035	201,91	0,639		
Dados completos					0.969	0.002

Prosseguiu-se com a transformação dos dados por Box-Cox (Figura 12). Contudo, embora tenha havido uma melhora com relação a ocorrência de pontos discrepantes e na distribuição dos dados, percebe-se que não foi suficiente para permitir que se prossiga com a análise de variância conjunta.



**Figura 12.** Gráficos de resíduos relacionados aos dados de trigo durum (*T. turgidum*) após transformação por Box-Cox para ambientes, genótipos e combinação genótipo e ambiente.

Considerando-se os resultados dos testes de hipóteses, apenas a combinação genótipo  $\times$  ambiente apresentou homogeneidade de variâncias. Contudo, os dados passaram a ter distribuição normal (Tabela 14).



**Tabela 14.** Testes de homogeneidade de variâncias para a variável rendimento de grãos (kg/ha) transformada, para cada fonte de variação e normalidade dos resíduos, considerando-se os dados de trigo durum (*T. turgidum*).

Fontes de Variação	Bartlett		Hartley		Shapiro-Wilk	
	$\chi^2$	Valor-p	Fmax	Valor-p	W	Valor-p
Ambiente	9,715	0,021	2,831	0,014		
Genótipo	34,595	0,000	11,982	0,010		
Gen x Amb	57,147	0,148	155,75	0,730		
Dados completos					0,984	0,102

### 4.3. Análise AMMI dos dados transformados de trigo comum (*T. aestivum*)

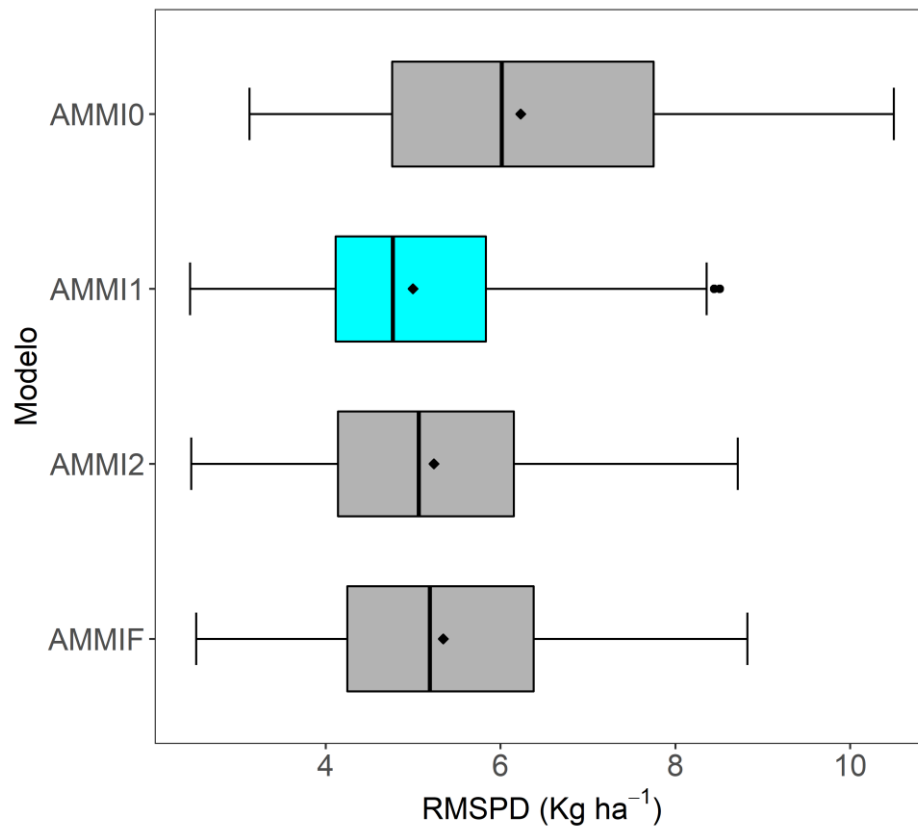
Considerando-se a tabela da análise de variância (Tabela 15), percebe-se que ambiente, bloco dentro de ambiente e genótipo dentro de ambiente foram significativos ao nível de 0,1% de probabilidade. Além disso, utilizando-se o teste F aplicado segundo a regra de Gollob, o modelo a ser selecionado é o AMMI1, que captura 89,6% da  $SQ_{G \times E}$ , sendo os demais eixos considerados como o ruído.

**Tabela 15.** Análise de variância para os dados transformados de rendimento de grãos em kg/ha de trigo comum (*T. aestivum*).

FV	GL	SQ	QM	F	Pr(>F)	Prop (%)	Acum (%)
Ambiente	3	5674,2	1891,4	90,4***	0,000		
Ambiente: bloco	8	167,3	20,9	13,8***	0,000		
Genótipo	11	306,6	27,9	18,3***	0,000		
Ambiente: Genótipo	33	221,4	6,7	4,4***	0,000		
IPC1	13	198,5	15,3	10,1	0,000***	89,6	89,6
IPC2	11	17,1	1,6	1,0	0,436 <sup>ns</sup>	7,7	97,3
IPC3	9	5,9	0,7	0,4	0,916 <sup>ns</sup>	2,7	100
Resíduo	88	133,7	1,5				
Total	176	6724,6	38,2				

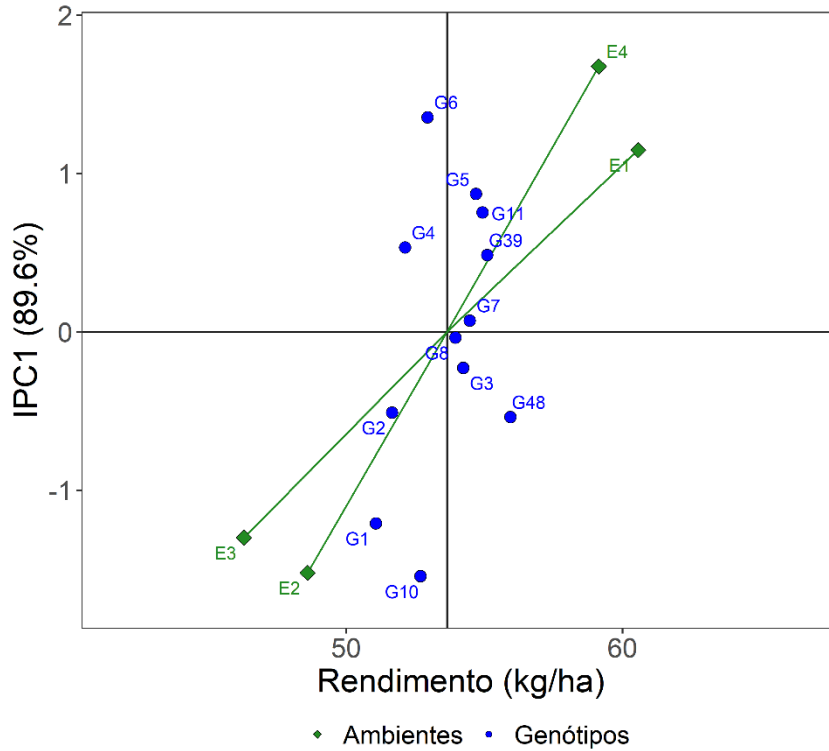
\*Significativo a 5% de probabilidade. \*\*\*Significativo a 0,1% de probabilidade. <sup>ns</sup>Não significativo.

Utilizando validação cruzada, conclui-se também que o modelo AMMI1 é o modelo a ser selecionado, considerando-se que este foi o que obteve o menor RSMPD (Figura 13).



**Figura 13.** Raíz cuadrada do quadrado médio da diferença de predição para os dados transformados de trigo comum (*T. aestivum*) e ajustados para a escala original dos dados.

Os genótipos G3, G7 e G8 foram os que menos contribuíram para a interação, sendo, portanto, os mais estáveis. Os genótipos G6 e G10, embora tenham tido uma produção média próxima da média geral, foram os mais instáveis. Os ambientes E1 e E4 são positivamente associados entre si, mas negativamente correlacionados com E2 e E3 (Figura 14).



**Figura 14.** Biplot AMMI1 (IPC1) *versus* rendimento médio (kg/ha) de ambientes e genótipos.

Conforme já se havia percebido na figura anterior, G7 e G8 podem ser considerados os mais estáveis e adaptáveis, considerando-se que estão mais ao centro do gráfico, mas a estabilidade do genótipo 3 não é confirmada. Além disso, G6 e G10 são confirmados como os menos estáveis e adaptáveis (Figura 15). Aqui, embora E1 e E2 sejam não correlacionados, vale ressaltar que a maior parte da variação explicada e significativa dos dados foi retida no modelo AMMI1.

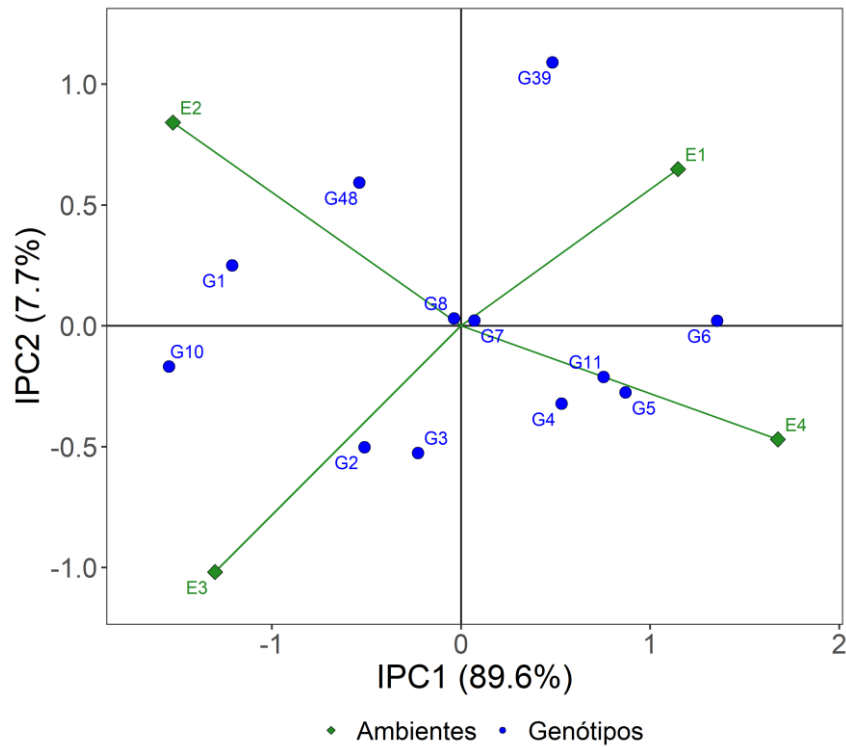


Figura 15. Biplot AMMI2 (IPC1 versus IPC2) dos dados transformados de trigo comum (*T. aestivum*).

#### 4.4. Aplicação dos modelos AMMI usando reamostragem

##### 4.4.1. Dados completos

Considerando-se os resultados do valor-p do bootstrap não paramétrico e da permutação, o modelo a ser selecionado é aquele que possui apenas os dois primeiros termos multiplicativos, que somados explicam 89,1% de toda a variação existente no conjunto de dados (Tabela 16).

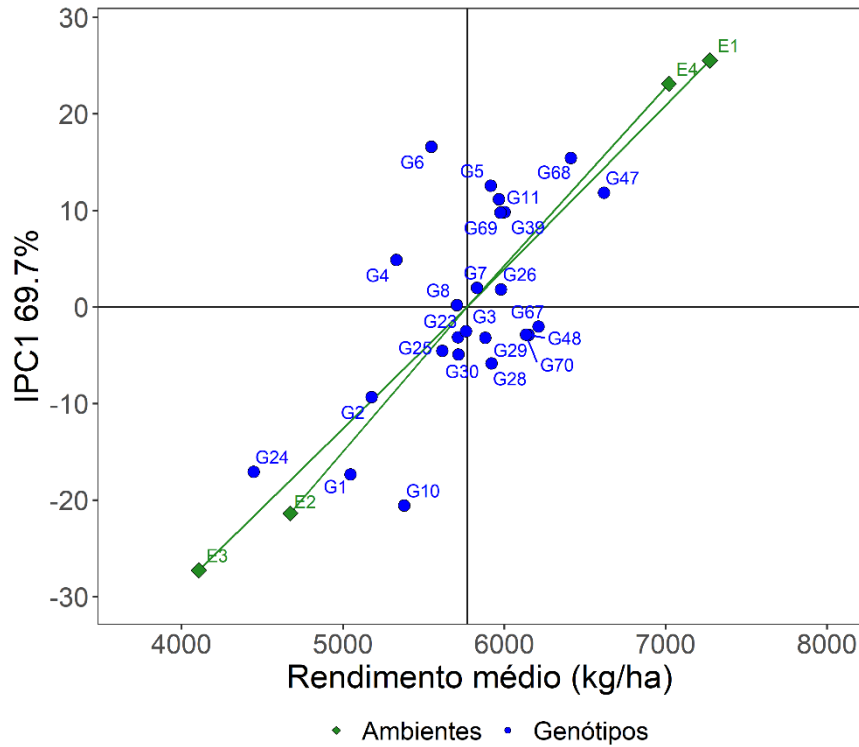
Tabela 16. Estatística  $Tr$  e valores-p calculados utilizando o método de *bootstrap* não-paramétrico e permutação para os dados completos.

Termo $K + 1$	$\hat{\lambda}_{K+1}$	Soma de Quadrados			$Tr^a$	Bootstrap não paramétrico <sup>b</sup>	Permutação <sup>b</sup>
		$\hat{\lambda}_{K+1}^2 B$	Acumulada	%*			
IPC1	2386,5	17086042	17086042	69,7	0,308	0,000	0,000
IPC2	1259,7	4760499	21846541	19,4	0,086	0,007	0,008
IPC3	940,0	2650605	24497146	10,8	0,048	0,158	0,151

\*Variação percentual explicada por um termo multiplicativo, ou seja, o valor singular ao quadrado dividido pela soma de todos os valores singulares ao quadrado. <sup>a</sup>Os testes assumem  $H_0: k = k$  contra  $H_1: k = k + 1$ . <sup>b</sup>Calculado em 5000 amostras de bootstrap.  $B$ : corresponde ao número de blocos.

Os ambientes um e quatro estão positivamente associados e, negativamente associados com os ambientes três e quatro, no entanto, todos tiveram alta contribuição para a interação. Os genótipos que menos contribuíram para a interação foram: G8, G7, G26, G3, G67, G70, G23 e

G48. Embora G47 seja altamente produtivo, ele também contribuiu muito para a interação, assim como G68. Estes dois últimos genótipos tiveram melhor desempenho nos ambientes E1 e E4 (Figura 16).



**Figura 16.** Biplot AMMI1 (IPC1) versus rendimento médio (kg/ha) dos dados completos de trigo, considerando-se as espécies *T. aestivum* e *T. durum*.

Plotando-se o segundo componente, verifica-se que os genótipos G8, G3, G7, G70, G23, G26, G4 e G29 estão localizados mais ao centro do gráfico, sendo, portanto, considerados os mais estáveis. O genótipo G48 que aparentemente era estável no gráfico anterior, agora aparece mais distante do centro da figura, não confirmando a sua aparente estabilidade apresentada na figura anterior. O ambiente um continua sendo negativamente correlacionado com o ambiente três e, de igual modo, os ambientes E2 e E3 possuem correlação negativa. G39 foi mais adaptável a E1, assim como, G47 e G69. Em E2, G10 foi o mais adaptável, em E3 foi G24 e, em E4 o G6 (Tabela 17).

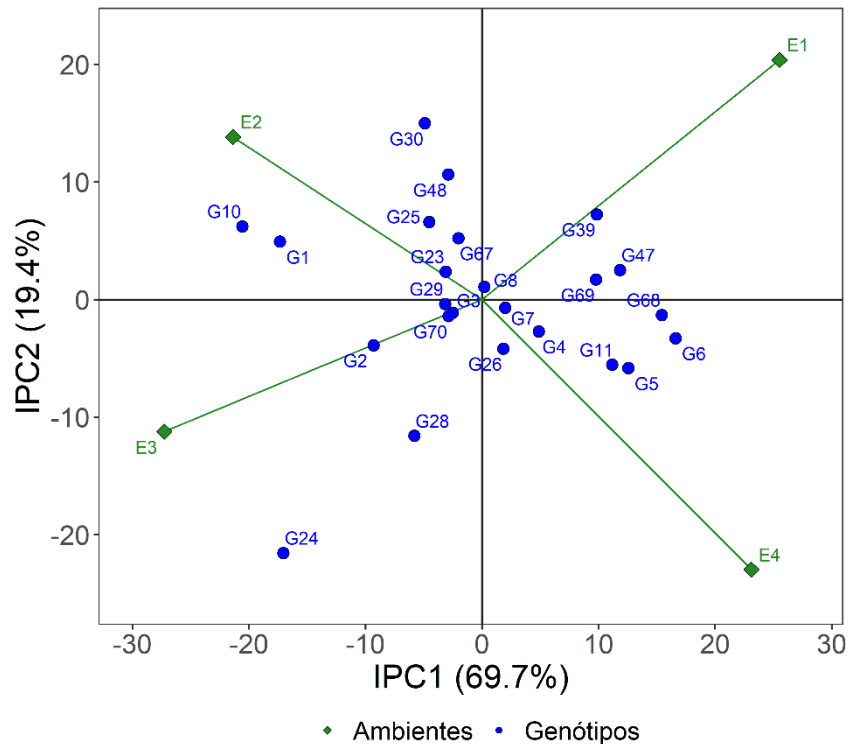


Figura 17. Biplot AMMI2 (IPC1 versus IPC2) para os dados completos, considerando-se as espécies *T. aestivum* e *T. durum*.

#### 4.4.2. Trigo durum (*T. turgidum*)

De acordo com os resultados dos testes para os três primeiros termos multiplicativos, utilizando-se o valor-p, pode-se observar que tanto o Bootstrap não paramétrico quanto a permutação sugerem que o modelo AMMI com os dois primeiros termos multiplicativos são apropriados para o conjunto de dados (Tabela 17).

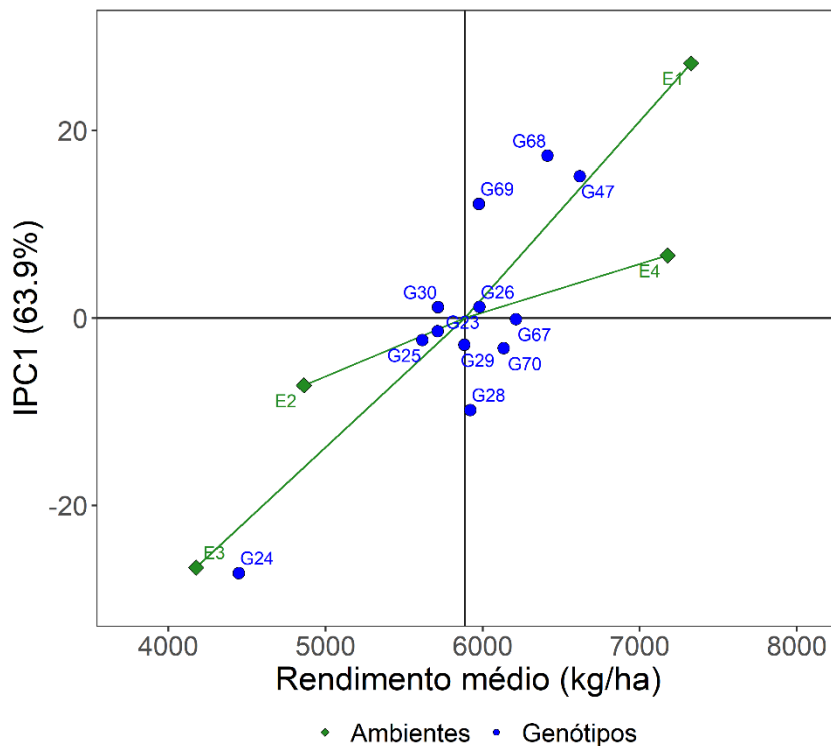
Tabela 17. Estatística  $Tr$  e valores-p calculados utilizando o método de *bootstrap* não-paramétrico e permutação para os dados de trigo

Termo $K + 1$	$\hat{\lambda}_{K+1}$	Soma de Quadrados			$Tr^a$	Bootstrap não paramétrico <sup>b</sup>	Permutação <sup>b</sup>
		$\hat{\lambda}_{K+1}^2 B$	Acumulada	%*			
IPC1	1543,2	7144516	7144516	63,9	0,214	0,000	0,000
IPC2	946,7	2688488	9833004	24,0	0,081	0,059	0,065
IPC3	671,3	1351908	11184912	12,1	0,040	0,210	0,213

\*Variação percentual explicada por um termo multiplicativo, ou seja, o valor singular ao quadrado dividido pela soma de todos os valores singulares ao quadrado. <sup>a</sup>Os testes assumem  $H_0: k = k$  contra  $H_1: k = k + 1$ . <sup>b</sup>Calculado em 5000 amostras de bootstrap.  $B$ : corresponde ao número de blocos

Considerando-se o IPC1, é possível observar que os ambientes um e quatro são positivamente associados entre si, mas ambos são negativamente associados com os ambientes dois e três. Os genótipos que menos contribuíram para a interação foram: G67, G26, G30, G23 e G25. À exemplo da análise AMMI para os dados completos, os genótipos G68 e G47, apesar de serem muito produtivos, são muito influenciados pela variação do ambiente. Contudo, aqui estes

genótipos são muito mais favorecidos pelo ambiente 1. Uma questão a se observar é que os ambientes E1 e E4 ainda possuem uma correlação positiva, porém, menor do que a observada ao se considerar os dados completos (Figura 18).



**Figura 18.** Biplot AMMI1 (IPC1) versus rendimento médio (kg/ha) para os dados de trigo durum (*T. turgidum*).

Embora o primeiro componente seja suficiente para explicar significativamente a maior parte da variação existente no conjunto de dados, decidiu-se considerar os dois primeiros componentes principais para avaliar se as mesmas interpretações se mantêm. Pode-se perceber que os genótipos G70 e G29 são os mais estáveis e o G23 agora aparece um pouco mais distante do centro do gráfico. É evidente que as conclusões desta segunda figura devem ser tomadas com cautela, considerando-se que IPC1 é o que tem a maior parte da variação dos dados, bem como, foi o único termo significativo apontado pela análise. Algo a ser destacado também é a menor associação existente entre E1 e E2 (Figura 19).

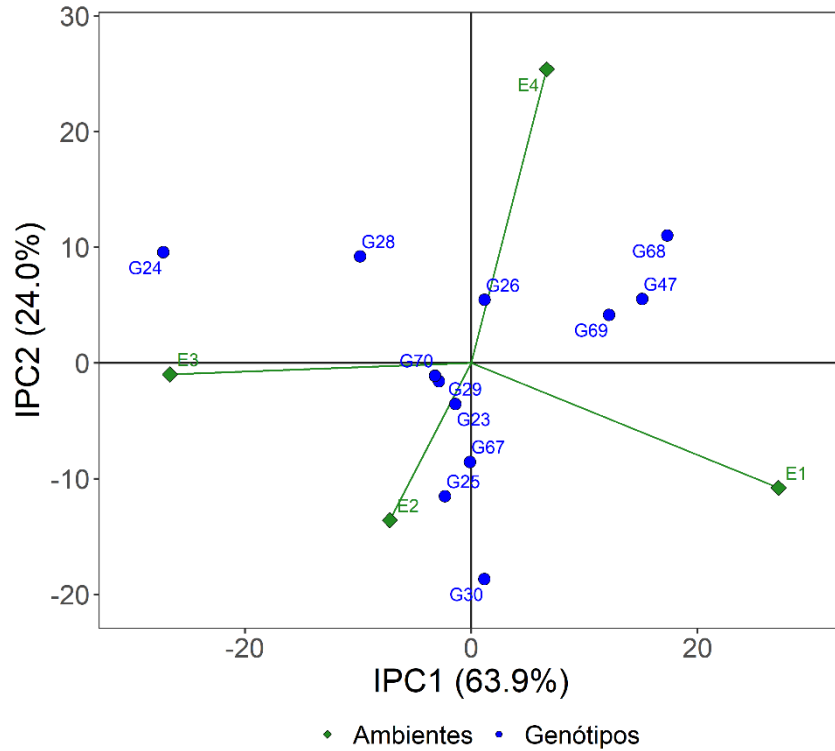


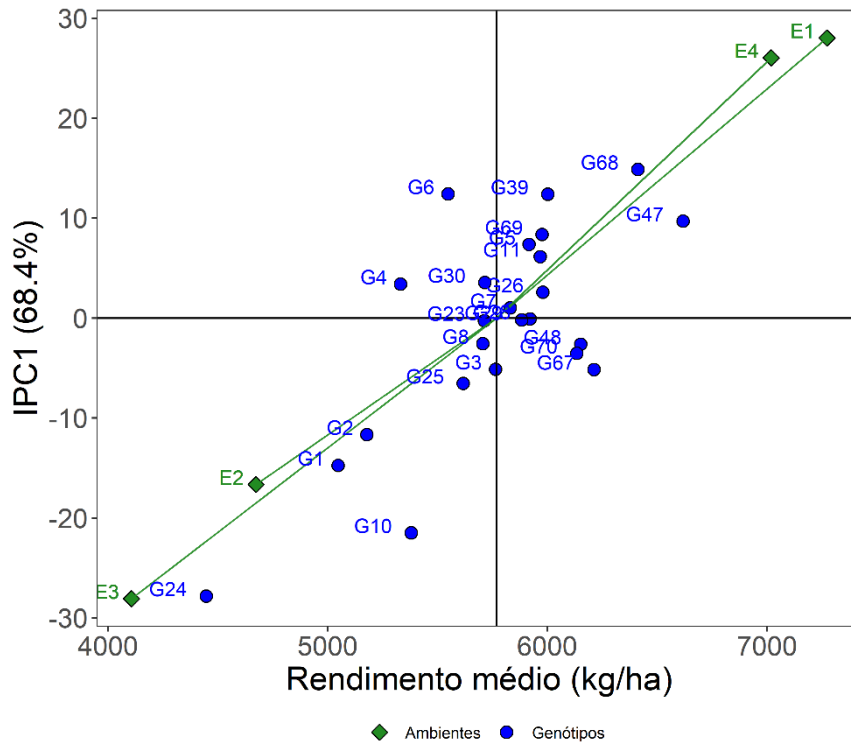
Figura 19. Biplot AMMI2 (IPC1 versus IPC2) dos dados de trigo durum (*T. turgidum*).

## 4.5. Aplicando o modelo R-AMMI

### 4.5.1. Dados completos

A proporção da variância explicada pelos componentes utilizando o algoritmo R-AMMI foi um pouco menor do que a análise AMMI tradicional, realizada a partir da DVS, com o IPC1 retendo 68,4% e o IPC2 16,33%. O gráfico gerado pela projeção do primeiro componente e das médias de genótipos e ambientes usando o algoritmo R-AMMI é muito semelhante aos resultados da análise AMMI tradicional, sendo os genótipos mais estáveis os mesmos já considerados anteriormente (Figura 20).





**Figura 20.** Biplot AMMI1 (IPC1) *versus* rendimento médio (kg/ha) dos dados completos utilizando o algoritmo R-AMMI.

Considerando-se o modelo R-AMMI para os dados completos com dois termos multiplicativos, percebe-se que embora a interpretação do relacionamento entre os ambientes seja semelhante ao que já se tinha anteriormente para os dados completos, aqui os ambientes E1 e E3 possuem uma correlação negativa muito mais expressiva. Além disso, o genótipo G48 que antes poderia ser considerado um genótipo instável, devido a sua grande contribuição para a interação, agora pode ser considerado um genótipo estável (Figura 21).

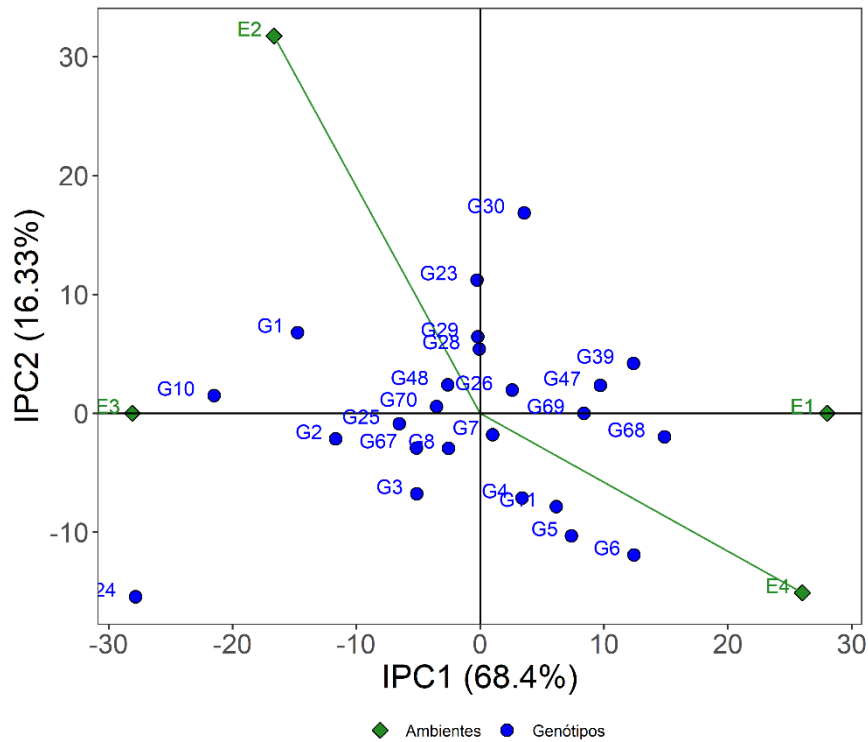
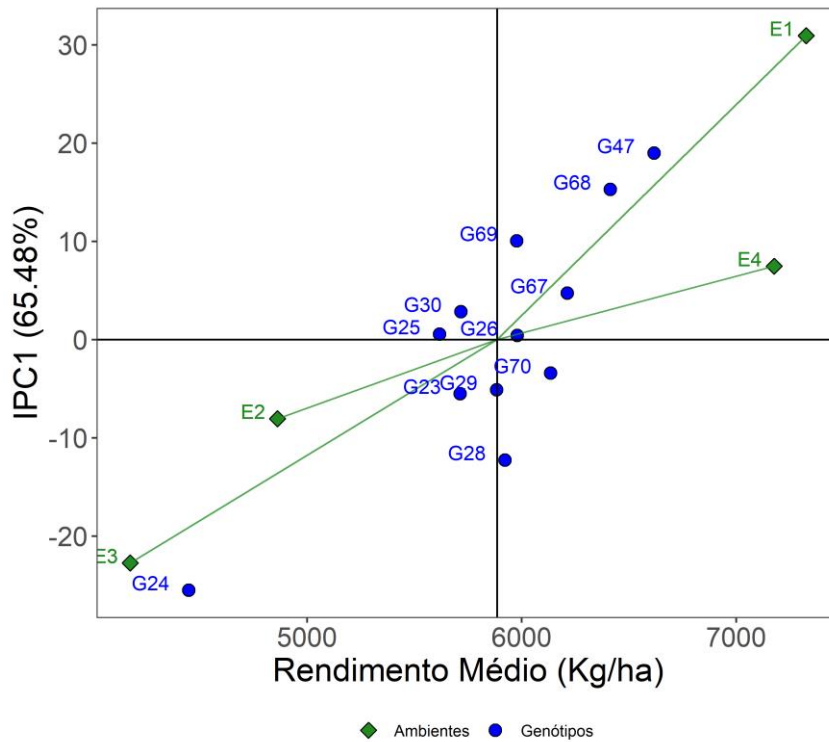


Figura 21. Biplot AMMI2 (IPC1 versus IPC2) dos dados completos usando o algoritmo R-AMMI.

#### 4.5.2. Trigo durum (*T. turgidum*)

A proporção da variância explicada pelos dois primeiros componentes da interação foi 87.9% utilizando o algoritmo R-AMMI, sendo 65,48% do IPC1 e 19,97% do IPC2. Embora a apresentação gráfica seja muito semelhante a apresentada anteriormente usando a abordagem tradicional de construção dos *biplots*, pode-se perceber uma diferença em relação aos genótipos considerados estáveis, como é o caso dos genótipos G67 e G23 que agora apresentam uma contribuição para a interação superior ao que se tinha observado anteriormente. De forma contrária, o genótipo G47 diminuiu a sua contribuição para a interação. Contudo, os genótipos G26, G25, G70 e G30 continuam sendo considerados estáveis (Figura 22).



**Figura 22.** Biplot AMMI1 (IPC1) versus rendimento médio (kg/ha) para a espécie de trigo durum (*T. turgidum*) usando o algoritmo R-AMMI.

Ao se considerar os dois primeiros termos multiplicativos, observa-se que há maiores diferenças em relação aos *biplots* construídos usando DVS. Percebe-se que os genótipos G30, G25 e G67 possuem menor contribuição para a interação. A estabilidade de G25 não é confirmada nesse segundo *biplot*, considerando-se como genótipos estáveis somente G23, G29 e G70 (Figura 23).

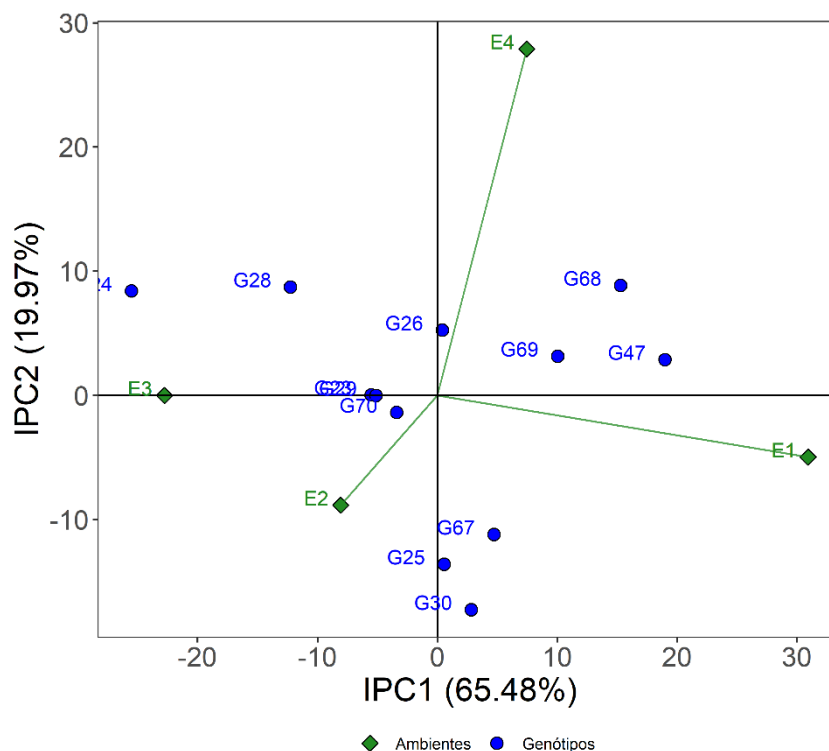


Figura 23. Biplot AMMI2 (IPC1 versus IPC2) para a espécie de trigo durum (*T. turgidum*) usando o algoritmo R-AMMI.

#### 4.6. Considerações sobre os resultados das análises

Embora alguns ambientes, quando analisados separadamente, não tenham apresentado pontos discrepantes, ao se analisar conjuntamente, todos os conjuntos de dados apresentaram essa característica. Observações discrepantes são valores que aparentemente não se encaixam quando comparados com os demais dados. No entanto, a sua exclusão deve ser realizada com cautela, considerando-se que aquela observação pode realmente expressar uma ocorrência real e não apenas um erro de medição ou digitação, particularmente quando o contexto atual caminha para a coleta de dados cada vez mais automatizada (NELSEN, 2023). Fazer uma boa análise residual é muito importante antes de aplicar o modelo AMMI, em função de ser possível identificar possíveis observações que podem influenciar na distribuição dos dados e na homogeneidade de variâncias residuais, além de aplicar os testes estatísticos adequados, a fim de se verificar se não será o caso de aplicar alternativas robustas ou transformações, considerando-se que a abordagem tradicional pode ser bastante sensível à presença de valores discrepantes (SA'DIYAH; HADI, 2023).

O sistema de plantio direto com irrigação reduzida, E4, teve um desempenho muito similar ao sistema de cultivo tradicional com irrigação completa, E1, em todas as análises. Sabe-se que em sistemas de PD, a perda de água para a atmosfera por evapotranspiração é menor,

ocorrendo uma otimização do teor de umidade do solo, regulando-se o uso da água durante os estágios vegetativos da planta (LIEBHARD et al., 2022; YIN et al., 2023). Isto pode explicar o desempenho produtivo semelhante entre esses ambientes, devido a manutenção dos estoques de água no solo em níveis adequados. Contudo, o excesso de irrigação parece ter prejudicado o ambiente sob PD. Isto pode ser explicado em parte devido a infiltração de água no solo em PD ser maior do que em preparo convencional (ACHARYA et al., 2019).

Assim como em outras culturas, o trigo também é fortemente influenciado por variações ambientais. Por exemplo, mudanças na quantidade de precipitação pluvial reduz a oferta de água às plantas, o que pode impactar em diversas variáveis, tais como em um menor índice de área foliar (IAF), número de grãos por espiga e peso de 1000 grãos, reduzindo consequentemente o rendimento final da cultura (POGGI et al., 2023). Em ambientes com menor quantidade de água disponível, espera-se que a interação genótipo  $\times$  ambiente seja mais complexa, reduzindo-se a capacidade estatística de identificar genótipos superiores (CHENU et al., 2011).

Em todas as análises, a dispersão dos pontos ou vetores relacionados aos ambientes contribuíram mais para a interação do que a dispersão de genótipos, se avaliados individualmente. Isto evidencia a importância do ambiente para o desempenho agrônômico de culturas e a necessidade de se utilizar técnicas estatísticas adequadas, como a análise AMMI, para analisar este tipo de fenômeno, e corrobora com resultados obtidos em outras pesquisas (BISHWAS; POUDEL; REGMI, 2021; NGAILO et al., 2019; SOOD et al., 2020; YOKOMIZO et al., 2013). As análises estatísticas tradicionais nem sempre são eficazes para analisar dados que podem ser influenciados não somente pelos efeitos principais, mas também, pela interação destes efeitos. Isto é, a ANOVA é capaz de verificar se os efeitos aditivos e multiplicativos são significativos ou não, mas não é capaz de decompor a interação. No caso da análise de componentes principais ou DVS, é possível analisar os efeitos multiplicativos, mas não os aditivos e, por fim, a análise de regressão somente poderá ser útil na análise dos termos da interação nos casos em que o padrão presente se ajustar a um modelo de regressão (ZOBEL; WRIGHT; GAUCH, 1988).

Honsdorf et al. (2018) analisaram os mesmos genótipos de trigo deste estudo, porém, avaliaram dados referentes aos anos/safras de 2009/10 a 2014/15. Segundo os autores, não há a necessidade de realização de programas de melhoramento genéticos separados para cada espécie, embora enfatizem que o progresso de seleção sob plantio direto ainda não tenha ficado claro. Além disso, os autores destacam que houve progresso reprodutivo de rendimento em todos os ambientes, mas principalmente nos que receberam irrigação total. Contudo, algo importante a se destacar é que, de acordo com os resultados das análises, é possível ter-se um ambiente altamente

produtivo usando o sistema de plantio direto, podendo-se alcançar produções similares à alcançada em cultivo convencional desde que haja o suprimento de água em condições adequadas às plantas.

A avaliação do potencial genético de um material de reprodução é de grande importância em um contexto de melhoramento genético. Ocorre que, como os ensaios são geralmente realizados em diferentes condições ambientais, artificiais ou não, o estudo da interação genótipo  $\times$  ambiente se torna bastante relevante para evitar recomendações equivocadas, ou mesmo fazer recomendações mais regionalizadas. A presença de heterogeneidade de variâncias entre ambientes é algo esperado, considerando-se que dados de campo são frequentemente influenciados por observações atípicas que podem influenciar no rendimento da cultura (LI et al., 2019; LI; TAO, 2022; XIN; TAO, 2019). Contudo, essa característica limita a aplicação da metodologia AMMI tradicional em função de não ser possível realizar a ANOVA e, conseqüentemente, escolher o número de termos multiplicativos significativos (GAUCH; GAUCH, 2013).

O número de termos multiplicativos a serem considerados no modelo é uma etapa crucial da Análise AMMI. Quando ocorre a violação das pressuposições da análise de variância, a escolha de quantos termos usar torna-se trabalhosa. Neste sentido, o uso de métodos baseados em reamostragem e livres de distribuições teóricas acabam se tornando bastante relevantes. O algoritmo proposto por Malik, Forkman e Piepho (2019), usando bootstrap não paramétrico e permutação se mostrou competitivo para testar o número de termos multiplicativos significativos em suas simulações. Segundo os autores, a metodologia proposta é superior ao teste *Fr* quando há heterogeneidade de variâncias, com ou sem normalidade de resíduos. Além disso, com um número crescente de *K* termos “verdadeiros”, a taxa de erro do tipo I diminui.

Os resultados obtidos por meio do modelo R-AMMI foram muito similares a análise AMMI clássica, com algumas excessões. Isto pode ter sido resultado possivelmente porque os dados não estão severamente afetados por observações discrepantes. De acordo com Rodrigues, Monteiro e Lourenço (2016), o modelo R-AMMI é muito semelhante a análise AMMI clássica quando os dados não estão contaminados com pontos discrepantes e, mesmo quando há a ocorrência deste cenário, ele ainda apresenta um desempenho satisfatório, com autovalores estimados menos tendenciosos, menor superestimação da proporção da variância explicada e biplots mais parcimoniosos.

O emprego do modelo robusto é indicado para os casos em que há interação entre os fatores, ocorrência de contaminação dos dados e não ser possível estabelecer pesos para observações atípicas. No caso da possibilidade de atribuir pesos às observações discrepantes,

recomenda-se o uso de um modelo AMMI ponderado, onde os pesos podem ser obtidos a priori (RODRIGUES et al., 2014). A aplicação do modelo AMMI robusto não é indicada nos casos em que os fatores não interagem entre si, podendo serem representados apenas pelos efeitos principais.

## 5. CONCLUSÕES

Os ambientes com canteiros convencionalmente arados com irrigação total e plantio direto com irrigação reduzida foram bastantes similares entre si, indicando que é possível ter um ambiente altamente produtivo com práticas conservacionistas do solo.

O sistema de plantio direto pode ser prejudicado no caso de excesso de irrigação, possivelmente devido a maior infiltração de água nesses ambientes, que podem resultar em estresse hídrico às plantas e, conseqüentemente, reduzir o rendimento de grãos de trigo.

Os algoritmos R-AMMI e o método baseado em reamostragem são importantes formas para contornar a situação de se trabalhar com dados contaminados com observações discrepantes, que implicam em não normalidade e heterogeneidade de variâncias dos resíduos.

Como sugestão para estudos futuros, recomenda-se combinar o algoritmo de reamostragem com o R-AMMI, aplicando-o diretamente nos resíduos da regressão robusta.





## REFERÊNCIAS

- ACHARYA, B. S. et al. Winter cover crops effect on soil moisture and soybean growth and yield under different tillage systems. **Soil and Tillage Research**, v. 195, p. 104430, 1 dez. 2019.
- ANGELINI, J.; PRUNELLO, M.; CERVIGNI, G. **geneticae: Statistical Tools for the Analysis of Multi Environment Agronomic Trials**. , 2022. Disponível em: <<https://cran.r-project.org/web/packages/geneticae/index.html>>. Acesso em: 4 abr. 2023
- BARTLETT, M. S. Properties of sufficiency and statistical tests. **Proceedings of the Royal Society of London. Series A - Mathematical and Physical Sciences**, v. 160, n. 901, p. 268–282, 18 maio 1937.
- BISHWAS, K. C.; POUDEL, M. R.; REGMI, D. AMMI and GGE biplot analysis of yield of different elite wheat line under terminal heat stress and irrigated environments. **Heliyon**, v. 7, n. 6, p. e07206, 1 jun. 2021.
- BOX, G. E. P.; COX, D. R. An Analysis of Transformations. **Journal of the Royal Statistical Society: Series B (Methodological)**, v. 26, n. 2, p. 211–243, 1 jul. 1964.
- BREUSCH, T. S.; PAGAN, A. R. A Simple Test for Heteroscedasticity and Random Coefficient Variation. **Econometrica**, v. 47, n. 5, p. 1287, set. 1979.
- CHENU, K. et al. Environment characterization as an aid to wheat improvement: interpreting genotype–environment interactions by modelling water-deficit patterns in North-Eastern Australia. **Journal of Experimental Botany**, v. 62, n. 6, p. 1743–1755, 1 mar. 2011.
- DIAS, C. T. D. S.; KRZANOWSKI, W. J. Model selection and cross validation in additive main effect and multiplicative interaction models. **Crop Science**, v. 43, n. 3, p. 865–873, 2003.
- DIAS, C. T. DOS S. **Métodos para escolha de componentes em modelo de efeito principal aditivo e interação multiplicativa (AMMI)**. Piracicaba - SP: Universidade de São Paulo, dez. 2005.
- DIAS, C. T. DOS S. et al. Introdução à metodologia AMMI. **Sigmae**, v. 2, n. 3, p. 38–56, 2013.
- DUARTE, J. B.; VENCOSKY, R. **Interação genótipos x ambientes: Uma introdução à análise AMMI**. Ribeirão Preto - SP: Sociedade Brasileira de Genética. Série Monografias, nº 9, 1999.
- EL HARCHE, S. et al. No-tillage and agroforestry decrease sediment loss from a hilly landscape in northern Morocco. **Catena**, v. 223, p. 106951, 1 abr. 2023.
- FISCHER, T. et al. Sixty years of irrigated wheat yield increase in the Yaqui Valley of Mexico: Past drivers, prospects and sustainability. **Field Crops Research**, v. 283, p. 108528, 1 jul. 2022.
- FISHER, R. A. XV.—The Correlation between Relatives on the Supposition of Mendelian Inheritance. **Earth and Environmental Science Transactions of The Royal Society of Edinburgh**, v. 52, n. 2, p. 399–433, 1919.
- GAUCH, H. G. **Model Selection and Validation for Yield Trials with Interaction**. [s.l: s.n.]. Disponível em: <<https://about.jstor.org/terms>>.
- GAUCH, H. G.; GAUCH, H. G. A Simple Protocol for AMMI Analysis of Yield Trials. **Crop Science**, v. 53, n. 5, p. 1860–1869, 1 set. 2013.
- GAUCH, H. G.; ZOBEL, R. W. Predictive and postdictive success of statistical analyses of yield trials\*. **Theoretical Applied of Genetics**, v. 76, p. 1–10, 1988.
- GAUCH, J. H. G. **Statistical analysis of regional yield trials : AMMI analysis of factorial designs**. New York: Elsevier Science, 1992.
- GAUCH, JR. H. G. Statistical Analysis of Yield Trials by AMMI and GGE. **Crop Science**, v. 46, n. 4, p. 1488–1500, 1 jul. 2006.

- GAUCH, JR. H. G. A Simple Protocol for AMMI Analysis of Yield Trials. **Crop Science**, v. 53, n. 5, p. 1860–1869, 1 set. 2013.
- GOLLOB, H. F. A statistical model which combines feature of factor analitic and analysis of variance techniques. **Psychometrika**, v. 33, n. 1, p. 73–115, 1968.
- GOOD, I. J. Some applications of the aingular decomposition of a matrix. **Technometrics**, v. 11, n. 4, p. 823–831, nov. 1969.
- HARTLEY, H. O. THE MAXIMUM F-RATIO AS A SHORT-CUT TEST FOR HETEROGENEITY OP VARIANCE. **Biometrika**, v. 37, n. 3–4, p. 308–312, 1 dez. 1950.
- HAWKINS, D. M.; LIU, L.; YOUNG, S. S. **Robust Singular Value Decomposition**. [s.l: s.n.]. Disponível em: <www.niss.org>.
- HONGYU, K. et al. Statistical analysis of yield trials by AMMI analysis of genotype  $\times$  environment interaction. **Biometrical Letters**, v. 51, n. 2, p. 89–102, 1 dez. 2014.
- HONSDORF, N. et al. Genotype by tillage interaction and performance progress for bread and durum wheat genotypes on irrigated raised beds. **Field Crops Research**, v. 216, p. 42–52, 1 fev. 2018.
- HONSDORF, N. et al. Dataset of historic and modern bread and durum wheat cultivar performance under conventional and reduced tillage with full and reduced irrigation. **Data in Brief**, v. 43, p. 108439, 1 ago. 2022.
- HUBER, P. J. **Robust statistics**. Nova York: Wiley, 1981.
- KARL PEARSON, F. R. S. LIII. On lines and planes of closest fit to systems of points in space. <https://doi.org/10.1080/14786440109462720>, v. 2, n. 11, p. 559–572, nov. 1901.
- LI, F. et al. Genetic architecture of grain yield in bread wheat based on genome-wide association studies. **BMC Plant Biology**, v. 19, n. 1, p. 1–19, 29 abr. 2019.
- LI, Y.; TAO, F. Interactions of genotype, environment and management on wheat traits and grain yield variations in different climate zones across China. **Agricultural Systems**, v. 203, p. 103521, 1 dez. 2022.
- LIEBHARD, G. et al. Effects of tillage systems on soil water distribution, crop development, and evaporation and transpiration rates of soybean. **Agricultural Water Management**, v. 269, p. 107719, 1 jul. 2022.
- LIMA NETO, E. DE A.; RODRIGUES, P. C. Kernel robust singular value decomposition. **Expert Systems with Applications**, v. 211, p. 118555, 1 jan. 2023.
- LIU, W. et al. Effects of no-tillage and biologically-based organic fertilizer on soil arbuscular mycorrhizal fungal communities in winter wheat field. **Applied Soil Ecology**, v. 178, p. 104564, 1 out. 2022.
- MALIK, W. A. et al. Nonparametric Resampling Methods for Testing Multiplicative Terms in AMMI and GGE Models for Multienvironment Trials. **Crop Science**, v. 58, n. 2, p. 752–761, 1 mar. 2018.
- MALIK, W. A.; FORKMAN, J.; PIEPHO, H. P. Testing multiplicative terms in AMMI and GGE models for multienvironment trials with replicates. **Theoretical and Applied Genetics**, v. 132, n. 7, p. 2087–2096, 1 jul. 2019.
- MANDEL, J. A new analysis of variance model for non-additive data. **Technometrics**, v. 13, n. 1, p. 1–18, fev. 1971.
- NELSEN, T. C. Outliers. **Probability and Statistics for Cereals and Grains**, p. 59–64, 1 jan. 2023.
- NGAILO, S. et al. Genotype-by-environment interaction of newly-developed sweet potato genotypes for storage root yield, yield-related traits and resistance to sweet potato virus disease. **Heliyon**, v. 5, n. 3, p. e01448, 1 mar. 2019.

- OLIVOTO, T.; LÚCIO, A. D. C. metan: An R package for multi-environment trial analysis. **Methods in Ecology and Evolution**, v. 11, n. 6, p. 783–789, 1 jun. 2020.
- PIOVESAN, P.; ARAÚJO, L. B. DE; DIAS, C. T. DOS S. Validação cruzada com correção de autovalores e regressão isotônica nos modelos de efeitos principais aditivos e interação multiplicativa. **Ciência Rural**, v. 39, n. 4, p. 1018–1023, 2009.
- POGGI, G. M. et al. Environment-oriented selection criteria to overcome controversies in breeding for drought resistance in wheat. **Journal of Plant Physiology**, v. 280, p. 153895, 1 jan. 2023.
- QI, J. Y. et al. A 40 % paddy surface soil organic carbon increase after 5-year no-tillage is linked with shifts in soil bacterial composition and functions. **Science of The Total Environment**, v. 859, p. 160206, 10 fev. 2023.
- R CORE TEAM. **R: A Language and Environment for Statistical Computing**. Vienna, Austria, 2021. Disponível em: <<https://www.R-project.org/>>. Acesso em: 4 abr. 2023
- RODRIGUES, P. C. et al. A Weighted AMMI Algorithm to Study Genotype-by-Environment Interaction and QTL-by-Environment Interaction. **Crop Science**, v. 54, n. 4, p. 1555–1570, 1 jul. 2014.
- RODRIGUES, P. C.; MONTEIRO, A.; LOURENÇO, V. M. A robust AMMI model for the analysis of genotype-by-environment data. **Bioinformatics**, v. 32, n. 1, p. 58–66, 1 jan. 2016.
- SA'DIYAH, H.; HADI, A. F. Outlying Observation in Stability Analysis of Genotype: AMMI vs Huehn Method. **AIP Conference Proceedings**, v. 2583, 13 jan. 2023.
- SCAVO, A.; MAUROMICALE, G.; IERNA, A. Genotype × environment interactions of potato tuber quality characteristics by AMMI and GGE biplot analysis. **Scientia Horticulturae**, v. 310, p. 111750, 15 fev. 2023.
- SHAPIRO, S. S.; WILK, M. B. An Analysis of Variance Test for Normality (Complete Samples). **Biometrika**, v. 52, n. 3/4, p. 591, dez. 1965.
- SOOD, S. et al. BLUP and stability analysis of multi-environment trials of potato varieties in sub-tropical Indian conditions. **Heliyon**, v. 6, n. 11, p. e05525, 1 nov. 2020.
- USDA - UNITED STATES DEPARTMENT OF AGRICULTURE. **World Agricultural Production**. [s.l.: s.n.]. Disponível em: <<https://apps.fas.usda.gov/psdonline/circulars/production.pdf>>. Acesso em: 5 abr. 2023.
- VERHULST, N.; DECKERS, J.; GOVAERTS, B. **Classification of the soil at CIMMYT's experimental station in the Yaqui Valley near Ciudad Obregon, Sonora, Mexico**. Ciudad Obregon: [s.n.].
- WIESER, H.; KOEHLER, P.; SCHERF, K. A. **Wheat - An Exceptional Crop: Botanical Features, Chemistry, Utilization, Nutritional and Health Aspects**. [s.l.] Elsevier, 2020.
- XIN, Y.; TAO, F. Optimizing genotype-environment-management interactions to enhance productivity and eco-efficiency for wheat-maize rotation in the North China Plain. **Science of The Total Environment**, v. 654, p. 480–492, 1 mar. 2019.
- YANG, H. et al. Maize straw mulching with no-tillage increases fertile spike and grain yield of dryland wheat by regulating root-soil interaction and nitrogen nutrition. **Soil and Tillage Research**, v. 228, p. 105652, 1 abr. 2023.
- YIN, W. et al. No-tillage with straw mulching promotes wheat production via regulating soil drying-wetting status and reducing soil-air temperature variation at arid regions. **European Journal of Agronomy**, v. 145, p. 126778, 1 abr. 2023.
- YOKOMIZO, G. K. I. et al. Análise AMMI da produtividade de grãos em linhagens de soja selecionadas para resistência à ferrugem asiática. **Pesquisa Agropecuária Brasileira**, v. 48, n. 10, p. 1376–1384, out. 2013.

ZOBEL, R. W.; WRIGHT, M. J.; GAUCH, H. G. Statistical Analysis of a Yield Trial. **Agronomy Journal**, v. 80, n. 3, p. 388–393, 1 maio 1988.